

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2004, 09:56:03 ; Search time 6532 Seconds
(without alignments)

Title: US-09-242-772-116_COPY_480_1980
Perfect score: 1501
Sequence: 1 gatggccactgctcctctg.....gtttccatcaagcttttcaag 1501

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb.ba.*
2: gb.hcg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1501	100.0	7313	6	A69511
2	1501	100.0	7313	9	HSU65002
3	1499.4	99.9	1565	9	BC075047
4	1499.4	99.9	1565	9	BC075048
5	1265.8	84.3	1575	10	AF057366
6	1265.8	84.3	1575	10	AF057366
7	1260.4	84.0	142102	9	AC107952
8	1038	69.2	278377	2	AC129839
9	1031.6	68.7	182102	2	AC097274
10	1031.6	68.7	241148	10	AL807387
11	964.2	64.2	244955	2	AC123210
12	389.8	26.0	1820	9	BC023655
13	389.8	26.0	2335	6	CQ834146
14	389.8	26.0	2429	9	AK026936
15	389.8	26.0	4051	9	AF006005
16	389.8	26.0	5638	9	D83784
17	388.2	25.9	5267	10	AB051854
18	388.2	25.9	5474	10	AK129087
19	377.8	25.2	5643	6	CQ727766

20	361.8	24.1	1437	10	AF181262
21	356.2	23.7	161478	9	AC025538
22	351.6	23.4	1491	9	BC074814
23	351.6	23.4	2334	6	BD005470
24	351.6	23.4	2725	6	CQ716510
25	351.6	23.4	2828	6	CQ834144
26	351.6	23.4	2828	9	HS8635
27	351.6	23.4	3215	6	AX834095
28	351.6	23.4	3215	9	AK096331
29	351.6	23.4	3215	9	HS8311395
30	350	23.3	2815	6	AX329760
31	350	23.3	3162	9	AK091707
32	350	23.3	4833	9	HSU72621
33	349.8	23.3	145414	9	HS8329218
34	349.6	23.3	200329	10	AC078911
35	349.6	23.3	208614	10	AL807380
36	348	23.2	240510	2	AC134371
37	347	23.1	240336	2	AC094844
38	344.4	22.9	1874	5	AF186476
39	301.8	20.1	5028	10	RNU72620
40	295.8	19.7	2350	10	AF324471
41	294.4	19.6	2771	10	MMZINCPR
42	294.4	19.6	2790	6	BD005460
43	294.4	19.6	3732	10	MMZINCPR
44	294.4	19.6	3975	6	AX961848
45	294.4	19.6	3975	10	AF147785

ALIGNMENTS

RESULT 1
LOCUS A69511 7313 bp DNA linear PAT 07-MAY-1999
DEFINITION Sequence 1 from Patent WO9807748.
ACCESSION A69511
VERSION A69511.1 GI:4774166
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 7313)
AUTHORS
TITLE PIAG GENE FAMILY AND TUMORIGENESIS
JOURNAL Patent: WO 9807748-A 1 26-FEB-1998;
VEN WILLEM JAN MARIE VAN DE (BE)
FEATURES
Location/Qualifiers
1..7313
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
481..1983
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB42491.1"
/db_xref="GI:4774167"
/translation="MATVIPGLSEVRTQKVPSCRRKRGKTPKRNFPQCLCDKAFN
SVEKLKHLVSHSGTGERPYKCIQDCTKAFVSKYKLRHMRHATHSPKTHKCNCEKVFH
RDKLKLHLHDPNKEIFKEECGNNTKLGFRHLHAHAATSGDLTKVCLQTFE
STGVLEHLKSHAGSGGVKQKHCEDRRFYRKDRVRHVMVTHGRKDFLCQY
AQRFGRKHLTHMKSHNBLKVTEPFDLPFCNVSVPIKDELLPMSLPS
LLSKPFTNLQNLNTPFQSMQSGAHOMITTLPLGTCPIQMDTVHPHSLSKFY
PFSSTVAISIPKEQPLKGEISYLMELQGVSPSSQSSSSKSLGMDPQISLD
DGAGDLSLSSKSSISITPLNTPALDPSOLFPLNGPPNPLSVSGLSGMSYSQBEAH
SSVQLPQTQDLPQDPANTIGLSLHLSAFTSLSTSTTLPRFHAQF"

CDS

ORIGIN

Query Match 100.0%; Score 1501; DB 6; Length 7313;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGCCACTGCTCCTCGTGTGTTGTTCAGAGTAGAGATACCCAGAAAGTCCTTC 60


```

/codon_start=1
/product="zinc finger protein pLAG1"
/protein_id="AAC50995.1"
/db_xref="GI:1839160"
/translation="MATVIFGDLSEVRDTQKVPSGKKRGKETPKRKNFPQCLCDKAFN
SVEKLKHSYSHTGERPYKCIQDDTKAFVSKYLQRHMATHSPSEKTHKNCYCEKMPH
RKDHLKHLHLLDPNKFTKCEEGKNYNTKLGFKRHLAHLAAHTSGDTCVKCLQTFE
STGVLLKHLKSHAGSGGVEKKHQCEHCDDRPYTRKDVRHMAWVTGSKDLCQXC
AQRFGKDHITRPMKKSHNOBLKVKTEPVDLPDPTCNYSVPIKDLELLPMSLPSE
LLSKPFTNTLQNLNYPFQSMQSSGAHQMIITPLGMCPCPIDMTVHPSHHLSFKY
PFSSTVAISITPEKEQPLKAEISYLMELQGVFPSSQDQAASSKXGLGDDPQIGSD
DGAGDLSKSGSISIDPLNTPALDFOLFPIPLNGPPYNPPLSVSGLSGMSYQEEAH
SSVSQLPSTQDQLQDPANTITGLGSLHSLSAFTSSISTSTTLPRFHQAQ"
723..7313
/number=5

```

Query Match	Best Local Similarity	100.0%;	Score 1501;	DB 9;	Length 7313;			
Matches 1501;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	GATGGCACTGTCATTCCTGGTGATTTGTGAGAGTAAGAGATACCCAGAAAGTCCCTTC	60					
Db	480							
Qy	61	AGGAAAACGTAAGCGTGGTGAACCCAAACCAGAAAAACCTTCTGCGCACTGTGCA	120					
Db	540	AGGAAAACGTAAGCGTGGTGAACCCAAACCAGAAAAACCTTCTGCGCACTGTGCA	599					
Qy	121	CAGGCCCTTAAACAGTGTGAGAAATTAAAGGTTCACTCTCTACACAGGAGAGAG	180					
Db	600	CAGGCCCTTAAACAGTGTGAGAAATTAAAGGTTCACTCTCTACACAGGAGAGAG	659					
Qy	181	GCCTCAAGTGCATACAAACAGACTGCACCAAGGCCCTTGTCTAAGTCAAAATTACA	240					
Db	660	GCCTCAAGTGCATACAAACAGACTGCACCAAGGCCCTTGTCTAAGTCAAAATTACA	719					
Qy	241	AAGGCATCGCTACTCATCTCTGAGAAAAACCCCAAGTGTAAATTATGTGAAAAAT	300					
Db	720	AAGGCATCGCTACTCATCTCTGAGAAAAACCCCAAGTGTAAATTATGTGAAAAAT	779					
Qy	301	GTTTTACCGGAAAGATCATCTGAAGATACCTCCATACACAGACCCCTAACAGAGAC	360					
Db	780	GTTTTACCGGAAAGATCATCTGAAGATACCTCCATACACAGACCCCTAACAGAGAC	839					
Qy	361	GTTTAAGTGCGAAGAAATGTGCGAAGAACTACAATACCAAGCTTCGATTTAAACGTCACTT	420					
Db	840	GTTTAAGTGCGAAGAAATGTGCGAAGAACTACAATACCAAGCTTCGATTTAAACGTCACTT	899					
Qy	421	GGCCTTGCAATGCCGCAACAGTGGTGACCTCACCTGTAAGGTATGTTGCAAACTTTTGA	480					
Db	900	GGCCTTGCAATGCCGCAACAGTGGTGACCTCACCTGTAAGGTATGTTGCAAACTTTTGA	959					
Qy	481	AAGCAGCGAGTGCTTCTGAGCACCTTAAATCTCATGAGGCAAGTCGTCGTGGGGT	540					
Db	960	AAGCAGCGAGTGCTTCTGAGCACCTTAAATCTCATGAGGCAAGTCGTCGTGGGGT	1019					
Qy	541	TAAAGAAAAAACGACCACTGGGAGCAATGTGATCGCCGTTCTACACCCGAAAGGATGT	600					
Db	1020	TAAAGAAAAAACGACCACTGGGAGCAATGTGATCGCCGTTCTACACCCGAAAGGATGT	1079					
Qy	601	CCGAGAGACATGGTGGTGACACTCGAAGAAAGGACTTCCTCTGTCAGTATTGTGCACA	660					
Db	1080	CCGAGAGACATGGTGGTGACACTCGAAGAAAGGACTTCCTCTGTCAGTATTGTGCACA	1139					
Qy	661	GAGATTTGGGCGAAAGGATCACTGACTCGACATATGAAGAAGAGTCACAATCAAGAGCT	720					
Db	1140	GAGATTTGGGCGAAAGGATCACTGACTCGACATATGAAGAAGAGTCACAATCAAGAGCT	1199					
Qy	721	TCTGAAGGTCAAAACAGAACCACTGGATTTCTTGAACCAATTTACTGCAATGTGCTGT	780					
Db	1200	TCTGAAGGTCAAAACAGAACCACTGGATTTCTTGAACCAATTTACTGCAATGTGCTGT	1259					

Qy	781	GCCTATAAAGACGAGCTCCTTCGGGTGATGCTCTTACCTTCCAGTGAACATGTTTATCAAA	840
Db	1260	GCCTATAAAGACGAGCTCCTTCGGGTGATGCTCTTACCTTCCAGTGAACATGTTTATCAAA	1319
Qy	841	GCCATTGCAAAACACTTTGGCAGTTTAAACCTCTACAACACTCCAATTCAGTCCCATGAG	900
Db	1320	GCCATTGCAAAACACTTTGGCAGTTTAAACCTCTACAACACTCCAATTCAGTCCCATGAG	1379
Qy	901	CTCGGAGTCTCCACCAAAATGATCAACAATTTACCTTTGGGAATGATCGCCCAATAGA	960
Db	1380	CTCGGAGTCTCCACCAAAATGATCAACAATTTACCTTTGGGAATGATCGCCCAATAGA	1439
Qy	961	TATGACACTGTTTCATCCCTCTCACCACTTTCTTTCAAATATCCGTTTCAGTTCTACCTC	1020
Db	1440	TATGACACTGTTTCATCCCTCTCACCACTTTCTTTCAAATATCCGTTTCAGTTCTACCTC	1499
Qy	1021	ATATGCAATTTCTATTCTCTGAAAGAACAGCCATTAAAGGGGAAATTCAGAGTTTACCT	1080
Db	1500	ATATGCAATTTCTATTCTCTGAAAGAACAGCCATTAAAGGGGAAATTCAGAGTTTACCT	1559
Qy	1081	GATGGAGTTTACAAGTGGCGTGCCCTCTTCATCCCAAGATTTCTCAAGCATCGTCAATC	1140
Db	1560	GATGGAGTTTACAAGTGGCGTGCCCTCTTCATCCCAAGATTTCTCAAGCATCGTCAATC	1619
Qy	1141	TAACTAGGGTTGGATCCTCAGATTGGGTCCCTTAGATGATGGTGAGGAGACTCTCCCT	1200
Db	1620	TAACTAGGGTTGGATCCTCAGATTGGGTCCCTTAGATGATGGTGAGGAGACTCTCCCT	1679
Qy	1201	ATCCAAAGCTCTATCTCCATGATGACGCCCTTAAACACACAGCATTTGGATTTTCTCA	1260
Db	1680	ATCCAAAGCTCTATCTCCATGATGACGCCCTTAAACACACAGCATTTGGATTTTCTCA	1739
Qy	1261	GTTGTTTAATTTTCATACCTTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCT	1320
Db	1740	GTTGTTTAATTTTCATACCTTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCT	1799
Qy	1321	TGGAATGAGCTATTCACGGAAGAAGCACATTTCTGTTTCCAGCTCCCCACACAAC	1380
Db	1800	TGGAATGAGCTATTCACGGAAGAAGCACATTTCTGTTTCCAGCTCCCCACACAAC	1859
Qy	1381	ACAGGATCTTCAGGATCTTGAAAACATATAGGGCTTGGGTCTCTGCACTCACTGTGAGC	1440
Db	1860	ACAGGATCTTCAGGATCTTGAAAACATATAGGGCTTGGGTCTCTGCACTCACTGTGAGC	1919
Qy	1441	AGCTTTCCACGAGTTTAAACACAAAGTACACCCCTCCAGGTTTCCATCAAGCTTTTCA	1500
Db	1920	AGCTTTCCACGAGTTTAAACACAAAGTACACCCCTCCAGGTTTCCATCAAGCTTTTCA	1979
Qy	1501	G 1501	
Db	1980	G 1980	

RESULT	3
LOCUS	BC075047
DEFINITION	Homo sapiens pleiomorphic adenoma gene 1, mRNA (cDNA clone MGC:103961 IMAGE:30915361), complete cds.
ACCESSION	BC075047
VERSION	BC075047.2 GI:50960384
KEYWORDS	MGC.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (Bases 1 to 1565) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Dege,J.G., Klausner,R.D., Collins,F.S., Wagner,L.H., Sherman,C.M., Schuler,G.D., Hopkins,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Diatchenko,L., Moore,T., Max,S.I., Wang,J., Hsieh,F., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
AUTHORS	

DB
942 CCGGAGTCGGCCACCAATGCTGTGTTTCTT

```

/gene="PLAG1"
/codon_start=1
/product="pleiomorphic adenoma gene 1"
/protein_id="AAH5047.1"
/db_xref="GI:49902412"
/db_xref="LocusID:5324"
/db_xref="MIM:603026"
/translation="NATVIPGLSEVRDTQPSGKRRGETPKQNFPOQLCDKAFN
SVKDLKLVHYSITGERTPKYCCQDCTAFVSQYKLQRHMAISPEKTHCNYCEKMFH
RKDLKLNHSTHIDPNKFTKEECEGKNYNTKLGPKRHLALHAATSGDLTCKVCLQTFE
STGVLLFLHLKSHAGKSSGVEKKGHCDDRRFYFTCDVRHNVVHTGKQPLCOYQ
AQRFGSKDLTHRMKSHNQELLKVTPEVDLPDFTTCNVSPVPIKELLPLNPSPSE

```

REMARK
COMMENT

cps

```
QY 961 TATGGACACTGTTTCATCCCTCTCACACACCTTCTTTCCTCAATATCCGTTCACTTCACTC 1020
Db |||
QY 1002 TATGGACACTGTTTCATCCCTCTCACACACCTTCTTTCCTCAATATCCGTTCACTTCACTC 1061
Db |||
QY 1021 ATATGCAATTTCTATTCCTGAAAGAAAGACAGCCATTAAAGGGGGAATAGAGAGTTACCT 1080
Db |||
QY 1062 ATATGCAATTTCTATTCCTGAAAGAAAGACAGCCATTAAAGGGGGAATAGAGAGTTACCT 1121
QY 1081 GATGGAGTTACAGGTGGGCGTGCCTCTTCATCCCAAGATTCTCAAGCATCGTCATCATC 1140
Db |||
QY 1122 GATGGAGTTACAGGTGGGCGTGCCTCTTCATCCCAAGATTCTCAAGCATCGTCATCATC 1181
QY 1141 TAAGCTAGGTTGGATTCCTCAGATTGGTCCCTAGATGATGGTGCAGGAGACCTTCCCT 1200
Db |||
QY 1182 TAAGCTAGGTTGGATTCCTCAGATTGGTCCCTAGATGATGGTGCAGGAGACCTTCCCT 1241
QY 1201 ATCCAAAAGCTCTATCTCCATCAGTGACGCCCTTAACACACAGCATTCGATTTTCTCA 1260
Db |||
QY 1242 ATCCAAAAGCTCTATCTCCATCAGTGACGCCCTTAACACACAGCATTCGATTTTCTCA 1301
QY 1261 GTTGTGTTAATTCATACCTTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCCT 1320
Db |||
QY 1302 GTTGTGTTAATTCATACCTTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCCT 1361
QY 1321 TGGATGAGCTATTCAGGAGAGACCATTTCTGTTCCAGTCCGCCACACAAAC 1380
Db |||
QY 1362 TGGATGAGCTATTCAGGAGAGACCATTTCTGTTCCAGTCCGCCACACAAAC 1421
QY 1381 ACAGGATCTTCAGGATCTGCAAAACACTATAGGCTTGCGTCTGCTGCACTCACTGTCAGC 1440
Db |||
QY 1422 ACAGGATCTTCAGGATCTGCAAAACACTATAGGCTTGCGTCTGCTGCACTCACTGTCAGC 1481
QY 1441 AGCTTTACAGAGCTTTAAAGCAAGTACCACCTCCAGCTTCCATCAAGCTTTTCA 1500
Db |||
QY 1482 AGCTTTACAGAGCTTTAAAGCAAGTACCACCTCCAGCTTCCATCAAGCTTTTCA 1541
QY 1501 G 1501
Db 1542 G 1542
```

```
RESULT 4
BC075048
LOCUS
DEFINITION
Homo sapiens pleiomorphic adenoma gene 1, mRNA linear PRI 04-AUG-2004
MGC:104123 IMAGE:30915577, complete cds.
ACCESSION
BC075048
VERSION
BC075048.2 GI:50960210
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1565)
Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Strausberg,R.L., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Sheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smialowski,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
```

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1565)
Director MGC Project.
Direct Submission
Submitted (25-JUN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 4, 2004 this sequence version replaced gi:49902222.
Contact: MGC help desk
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: Genome Sequence Centre, British Columbia Cancer Center
cDNA Library Preparation: British Columbia Cancer Research Center
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabbu, Parvaneh Saeedi, JR Santos, Angeliq Schnerch, Ursula Skalska, Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacques Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRBU Plate: 5 Row: C Column: 2.
Location/Qualifiers

1. 1565
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="MGC:104123 IMAGE:30915577"
/tissue_type="Brain, PCR rescued clones"
/clone_lib="NIH-MGC_264"
/lab_host="DH10B"
/notes="Vector: pCR4 Topo TA"
1. 1565
/gene="PLAG1"
/note="synonyms: SGPA, PSA"
/db_xref="LocusID:5324"
/db_xref="MIM:603026"
43. 1545
/genes="PLAG1"
/codon_start=1
/product="pleiomorphic adenoma gene 1"
/protein_id="AAH75048.1"
/db_xref="GI:49902223"
/db_xref="LocusID:5324"
/translation="MATVTPGDLSEVDTQKVPKGRKRGKTKPRKNFPQCLCDKAFN
SVEKLKHSYSHTGRTGPKYKIQDCTKAFVSKYKLRHWAHSPEKHKNCYCEKMFH
RDKLKNLHLDHPKRTFKCECKNNTKLGFKRHLAHLAHAATSGDILTKVCLQTFE
STGVLKHLHSHAGSKSGGVKKEGKCECHDRFTRKDVRRHVMVTHGRKDFLQCYC
AQGRKDKHLTRHMKSHNQLLVKTPVDFLDPDFTCNVSVPIKDELLPWSLPSSE
LLSKPTNLQNLNTPFQSMQSSGAHQMITLPLGTCPLDMDTVPHSHLSFKY
PFSSTSYAISPEKEQPLKGEISYLMELQGVSSSDSQSSSSKGLGDPQIGSLD
DGAGLSLSSKSSISISDPLNTPALDFSQLFNPPLNGPPNPLSVSLGMSYSQBEAH
SSVSQLPQTDQDLPANTIGLSLHLSAAFTSLSTSTLTPRHQAQ"

gene

CDS

ORIGIN

Query Match 99.9%; Score 1499.4; DB 9; Length 1565;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 CATGCCACTGTCATTCTGCTGATTTGTTCAGAGTAAGAGATACCCAGAGAAGTCCCTTC 60

Db	1122	GATGGAGTTACAAAGTGGCGTGCCTCTTTCATCCCAAGATTCTCAAGCATCGTCATCATC	1181
Qy	1141	TAAGCTAGGGTTGGATCCCTCAGANTGGTCCCTAGATGATGGTGCAGGACCTCTCCCT	1200
Db	1182	TAAGCTAGGGTTGGATCCCTCAGANTGGTCCCTAGATGATGGTGCAGGACCTCTCCCT	1241
Qy	1201	ATCCAAAAGCTCTATCTCCATCAGTGAACCCCTTAAACACACACACAGCATTTGATTTCTCA	1260
Db	1242	ATCCAAAAGCTCTATCTCCATCAGTGAACCCCTTAAACACACACACAGCATTTGATTTCTCA	1301
Qy	1261	GTTCGTTTAAATTTTCACTCTTTAAATGGTCTCCCTTAAATCTCTATCAGTGGGAGCCT	1320
Db	1302	GTTCGTTTAAATTTTCACTCTTTAAATGGTCTCCCTTAAATCTCTATCAGTGGGAGCCT	1361
Qy	1321	TGGAATGAGCTATTTCCAGGAAGAGACATTTCTCTCTTTCCAGCTCCCCACACAAAC	1380
Db	1362	TGGAATGAGCTATTTCCAGGAAGAGACATTTCTCTCTTTCCAGCTCCCCACACAAAC	1421
Qy	1381	ACAGATCTTCAGGATCTCTCAACACATATAGGGCTTGGGCTCTGCACTCATCTGACG	1440
Db	1422	ACAGATCTTCAGGATCTCTCAACACATATAGGGCTTGGGCTCTGCACTCATCTGACG	1481
Qy	1441	AGCTTTCCAGGACGTTTAAAGCACAACTACCAAGTACCACTCCCAAGCTTTTCA	1500
Db	1482	AGCTTTCCAGGACGTTTAAAGCACAACTACCAAGTACCACTCCCAAGCTTTTCA	1541
Qy	1501	G 1501	
Db	1542	G 1542	
RESULT 5			
AF057366 1575 bp mRNA linear ROD 01-JAN-2000			
LOCUS Mus musculus C2H2 zinc finger protein FLAG1 (Plagl) mRNA, complete			
DEFINITION cds.			
ACCESSION	AF057366	GI:6650194	
VERSION	AF057366.1		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 1575)		
AUTHORS	Kas, K., Voz, M.L., Roljer, E., Astrom, A.K., Meyen, E., Stenman, G. and Van de Ven, W.J.		
TITLE	Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations		
JOURNAL	Nat. Genet. 15 (2), 170-174 (1997)		
MEDLINE	97172974		
PUBMED	9020842		
REFERENCE	2 (bases 1 to 1575)		
AUTHORS	Hensen, K., Voz, M.L., Van de Ven, W.J.M. and Kas, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-APR-1998) Laboratory of Molecular Oncology, Center of Human Genetics, Herestraat 49, Leuven 3000, Belgium		
FEATURES	Location/Qualifiers		
source	1..1575		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="129/SV"		
	/db_xref="taxon:10090"		
	/map="t(3;8)(p21;q12)"		
	/tissue_type="pleiomorphic adenoma"		
gene	1..1575		
	/gene="Plagl"		
	/note="synonym: pleiomorphic adenoma gene 1"		
CDS	21..1520		
	/gene="Plagl"		
	/codon_start=1		
	/product="C2H2 zinc finger protein FLAG1"		
	/protein_id="AAF21762.1"		
Db	42	GATGGCAGCTGTATTCTCTGGTGAATTTGTCAGAGTAAGAGATACCCAGAAAGTCCCTTC	101
Qy	61	AGGGAACAGTAAGCGTGGTGAACCAACCAAGAAAAATTTTCCTTGGCAACTGTGTGA	120
Db	102	AGGGAACAGTAAGCGTGGTGAACCAACCAAGAAAAATTTTCCTTGGCAACTGTGTGA	161
Qy	121	CAAGGCGCTTTAAACAGTGTGAGAAATTAAGAGTTTCACTCTCTACCTCAGCAGAGAGAG	180
Db	162	CAAGGCGCTTTAAACAGTGTGAGAAATTAAGAGTTTCACTCTCTACCTCAGCAGAGAGAG	221
Qy	181	GCCCTACAGTGCATACCAAGAGTGCACCAAGCGCTTTTCTTCTAAGTACAAATTACA	240
Db	222	GCCCTACAGTGCATACCAAGAGTGCACCAAGCGCTTTTCTTCTAAGTACAAATTACA	281
Qy	241	AAGGCACATGCTACTCTATCTCTCTGAGAAACCCCAAGTGTATTTATTTGTGAGAAAT	300
Db	282	AAGGCACATGCTACTCTATCTCTCTGAGAAACCCCAAGTGTATTTATTTGTGAGAAAT	341
Qy	301	GTTCACCGGAAAGATCTATCTGAGAAATCACTCTCATACAGCAGCCTTAAACAAAGAGAC	360
Db	342	GTTCACCGGAAAGATCTATCTGAGAAATCACTCTCATACAGCAGCCTTAAACAAAGAGAC	401
Qy	361	GTTCAGTGGGAGATGTCGAGAAATCACTCTCATACAGCAGCCTTAAACAAAGAGAC	420
Db	402	GTTCAGTGGGAGATGTCGAGAAATCACTCTCATACAGCAGCCTTAAACAAAGAGAC	461
Qy	421	GCCCTTGATGCGGCAACAGTGGTGCACCTCACTGTAAGGTATGTTTCAAACTTTTGA	480
Db	462	GCCCTTGATGCGGCAACAGTGGTGCACCTCACTGTAAGGTATGTTTCAAACTTTTGA	521
Qy	481	AAGCAGCGGAGTCTTCTGAGACACCTTAAATCTCATGAGCAAGTCTGCTGGGCT	540
Db	522	AAGCAGCGGAGTCTTCTGAGACACCTTAAATCTCATGAGCAAGTCTGCTGGGCT	581
Qy	541	TAAAGAAAAAAGCAGCAGTGGCAACATTTGTGATCGCCGTTCTACACCCGAAAGATGT	600
Db	582	TAAAGAAAAAAGCAGCAGTGGCAACATTTGTGATCGCCGTTCTACACCCGAAAGATGT	641
Qy	601	CCGGAGACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTCAGTATTTGTCACA	660
Db	642	CCGGAGACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTCAGTATTTGTCACA	701
Qy	661	GAGATTGGCGGAAAGATCACTGACCTGCACATATGAAGAGAGTCAATCAAGAGCT	720
Db	702	GAGATTGGCGGAAAGATCACTGACCTGCACATATGAAGAGAGTCAATCAAGAGCT	761
Qy	721	TCTGAGGTCAAAACAGAACAGTGGATTTCTTTGACCCCAATTTACCTGCAATGTCTGT	780
Db	762	TCTGAGGTCAAAACAGAACAGTGGATTTCTTTGACCCCAATTTACCTGCAATGTCTGT	821
Qy	781	GCCTATAAAGACGAGCTCTTCCGGTGAATGCTTACCTTCCAGTGAAGTATCAAA	840
Db	822	GCCTATAAAGACGAGCTCTTCCGGTGAATGCTTACCTTCCAGTGAAGTATCAAA	881
Qy	841	GCCATTCACAAACACTTTGAGTAACTCTGACACTCCATTTGAGTCCATGCGAGAG	900
Db	882	GCCATTCACAAACACTTTGAGTAACTCTGACACTCCATTTGAGTCCATGCGAGAG	941
Qy	901	CTCGGATCTGCCACCAATGATCACAACCTTTACCTTTGGGAATGATGCCCAATAGA	960
Db	942	CTCGGATCTGCCACCAATGATCACAACCTTTACCTTTGGGAATGATGCCCAATAGA	1001
Qy	961	TATGACACTGTTATCCCTCTCAACAACCTTTACCTTTGGGAATGATGCCCAATAGA	1020
Db	1002	TATGACACTGTTATCCCTCTCAACAACCTTTACCTTTGGGAATGATGCCCAATAGA	1061
Qy	1021	ATATGCAATTTCTATCTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1080
Db	1062	ATATGCAATTTCTATCTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1121
Qy	1081	GATGGAGTTACAGGTGGGCTGCCCTCTTCATCCCAAGATTTCTCAAGATCGTCATCATC	1140

/db_xref="GI:6650195"		/translation="MATVPGDLSFVRDQKAPSGKRRKESKPRNPFQCLCDKAFN		860		GCCATTCAAAACACCTTTGAGATTAAACCTCTATAACACCTCCATTCAGTCCATGACAG		919		
SVKLVHFSHTGSRPYKCTHODCKAFVSKYKQLRHATHSPETHKNCYCEKMFH		RKDHLKHLHTDHPNETFKCEGKSYNFKLGFKEHLALHAATSDGLTKVCLOQNE		901	CTCGGAGCTGCCCAACCAATGATCAAACTTTACCTTTGGGAAATGACATGCCCAATAGA		960			
STGVLEHLKSHAGKSGGVKKHQCERRFYTRKDVRRHVVHTGRKDFLOCYC		AQRFGKDLTRHMLKSHQELLKVTEPVDLPDFTCNMSPVTKDELPLVMSLPSE		920	CTCTGGGTCTGCTCACCAATGATAAACTTTACCTTTGGGAAATGACATGCCCAATAGA		979			
LLSKPFTNLQNLVNTPFQSMQSSGSAHQMTTLPLGNTCPIDMDVHPHSLAFKC		PFSSSYAISIPKBPQKGEIESYLMELQGGAPSSQDPSASSKLGLEPQSGPDD		961	TATGGACACTGTTTCATCCCTCTCACACCTTTCTTTCAAATATCCGTTTCAGTTTCTACCTC		1020			
GAGDLSLSSSISIDPLSTPALDFQNFPIPLNGPPYNTLSVSLGMSYSQEBAS		SVSQPTQDQLQDPANTVGLSLSLSAFTSSLSSTLTPRFHQAFQ"		980	TATGGATGCTGTTTACCCCTCTCATCATCTTCTCTTCAAATGCCATTTCAGTTCTACCTC		1039			
ORIGIN				1021	ATATGCAATTTCTATTCTCTGAAAAGAACAGCCATTAAGAGGGGAAAATGAGATTACCT		1080			
Query Match		84.3%; Score 1265.8; DB 10; Length 1575;		1040	ATACGCAATCTCTATCTCTGAAAAGAACAGCCATTAAGAGGGGAAAATGAGATTACCT		1099			
Best Local Similarity		90.7%; Pred. No. 0;		1081	GATGAGTTTACAGGTGGCGTCCCTCTTCATCCCAAGATTTCTCAAGCATCTGCATCATC		1140			
Matches 1361; Conservative 0; Mismatches 137; Indels 3; Gaps 1;				1100	GATGAGCTTACAGGTGGTGACCATCTTCATCCAGGATTTCTCCAGCATCG---TCATC		1156			
QY	1	GATGCGCATGTCTATCTCTGTGATTGTTCAGAAAGTAAAGAGATACCCAGAAAGTCCCTTC	60	1141	TAAGCTAGGTTGGATCTCAGATTGGGTCCCTAGATGATGGTGAGGAGACCTCTCCCT		1200			
DB	20	GATGCCCATGTCTATCTCTGTGATTGTTCAGAAAGTAAAGAGATACCCAGAAAGTCCCTTC	79	1157	TAAGCTAGGTTAGAGCTCTCAGTGGGTGCGCAGATGATGGTGGGACCTCTCCCT		1216			
QY	61	AGGGAAGCTAAGCTGTGAAACCAACCAAGAAACCAAGAAACCAAGAAACCAAGAAAC	120	1201	ATCCAAAAGCTCTATCTCCATCAGTGAGTGGGTCCCTAGATGATGGTGAGGAGACCTCTCCCT		1260			
DB	80	AGGGAAGCTAAGCTGTGAAACCAACCAAGAAACCAAGAAACCAAGAAACCAAGAAAC	139	1217	GTCAAAGAGCTCTATCTCTATCAGTGACCCCTCAGCACAGCAGCATTTGGATTTCTCTCA		1276			
QY	121	CAAGGCTTTAAACAGTTGTGAGAAATTAAGGTTCACTCTCTACTCTCAACAGGAGAGAG	180	1261	GTTGTTTAAATTTATACCTTTAAATGGTCCCTCTATAATCTCTATCAGTGGGAGCCT		1320			
DB	140	CAAGGCTTTAAACAGTTGTGAGAAATTAAGGTTCACTCTCTACTCTCAACAGGAGAGAG	199	1277	GTTGTTTAAATTTATACCTTTAAATGGTCCCTCTATAATCTCTATCAGTGGGAGCCT		1336			
QY	181	GCCCTACAGTGCATACCAACAGACTGCACCAAGGCTTTGTTTCTTAAGTACAAATTA	240	1321	TGGATGAGCTATTCCAGGAGAAAGACATTTCTTGTTCAGAGTCCCAACAAAC		1380			
DB	200	GCCCTACAGTGCATACCAACAGACTGCACCAAGGCTTTGTTTCTTAAGTACAAATTA	259	1337	TGGATGAGCTATTCCAGGAGAAAGACATTTCTTGTTCAGAGTCCCAACAAAC		1396			
QY	241	AAGGCATGCTACTACTCTCTCTGAGAAACCAACCAAGTGAATTTATTTGAGAAAT	300	1381	ACAGGATCTTCAGGATCTCTGCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGC		1440			
DB	260	AAGGCATGCTACTACTCTCTCTGAGAAACCAACCAAGTGAATTTATTTGAGAAAT	319	1397	ACAGGATCTTCAGGATCTCTGCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGC		1456			
QY	301	GTTTTCAGGAAAGATCATCTGAAAGATCACTCCATACACACAGCCCTTAACAAAGAGAC	360	1441	AGCTTTCAGGATCTCTGCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGC		1500			
DB	320	GTTTTCAGGAAAGATCATCTGAAAGATCACTCCATACACACAGCCCTTAACAAAGAGAC	379	1457	AGGCTTTCAGGATCTCTGCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGC		1516			
QY	361	GTTTTCAGGAAAGATCATCTGAAAGATCACTCCATACACACAGCCCTTAACAAAGAGAC	420	1501	G 1501					
DB	380	CTTTAAATGGAAGAGTGTGGCAAGAGTCAATACAGCTTGGTTTAAAGCACTT	439	1517	G 1517					
QY	421	GGCTTTGATGCTGCAACAGTGTGACCTCACTGCAAGGATGTTTGAACCTTTGA	480	RESULT 6						
DB	440	GGCTTTGATGCTGCAACAGTGTGACCTCACTGCAAGGATGTTTGAACCTTTGA	499	AY574219	3161 bp mRNA linear ROD 12-APR-2004					
QY	481	AAGCAGGGAGTGTCTTCTGAGCAGCTTAATCTCATGAGGAGTGTCTGGTGGGT	540	LOCUS	Mus musculus pleomorphic adenoma of the salivary gland 1 (Plagl1)					
DB	500	GAGCAGAGTGTGCTCTCTAGAGCAGCTTAATCTCATGAGGAGTGTCTGGTGGGT	559	DEFINITION	mRNA, complete cds; alternatively spliced.					
QY	541	TAAAGAAAAGACACAGTGTGCAACATTTGTGATCGCGTCTTCTACACCGAAGATGT	600	ACCESSION	AY574219					
DB	560	GAAGGAGAAAAGACCAATGTGAGCACTGCGAAGCAGGTTCTACACCGGAGAGATGT	619	VERSION	AY574219.1					
QY	601	CCGAGACACATGTTGTCACACTGGAAGAAAGACTTCTCTGTGAGTATTGTGCACA	660	KEYWORDS	GI:46254398					
DB	620	CCGAGACACATGTTGTCACACTGGAAGAAAGACTTCTCTGTGAGTATTGTGCACA	679	SOURCE	Mus musculus (house mouse)					
QY	661	GAGATTGGGCGAAGAGTCACTGAGTGCATATGAAGAGAGTCAATCAAGAGCT	720	ORGANISM	Mus musculus					
DB	680	GAGATTGGGCGAAGAGTCACTGAGTGCATATGAAGAGAGTCAATCAAGAGCT	739	REFERENCE	1. .3161					
QY	721	TCTGAAGGTCAAAACAGACCAAGTGTGATTCTCTGACCCATTTACTGCAATGTCTGT	780	AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
DB	740	TCTGAAGGTCAAAACAGAACCAAGTGTGATTCTCTGACCCATTTACTGCAATGTCTGT	799	TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
QY	781	GCCTATAAAGCAGAGCTCTTCCGGTGTGATGTCCTTACCTTCCAGTGAACCTTATCAA	840	JOURNAL	1 (bases 1 to 3161)					
DB	800	GCCTATAAAGTGAACCTCTCCGGTGTGATGTCCTTACCTTCCAGTGAACCTTGTCAA	859	FEATURES	Castilla, L.H.					
QY	841	GCCATTCAAAACACTTTTGCACTTAACTCTTACAACTCACTCACTTCACTGATGATG	900	Location/Qualifiers	Submitted (16-MAR-2004) Program in Gene Function and Expression,					
					University of Massachusetts Medical School, 364 Plantation Street,					
					Worcester, MA 01605, USA					
					/organism="Mus musculus"					
					/mol_type="mRNA"					
					/strain="129/SvEv"					
					/db_xref="taxon:10090"					
					/chromosome="4"					

Query Match	84.3%;	Score 1265.8;	DB 10;	Length 316.7;
Best Local Similarity	90.7%;	Pred. No. 0;		
Matches 1361;	Conservative	0;	Mismatches 137;	Indels 3; Gaps 1;
QY	1	GATGCCACTGTCAATTCTCGTGTAATTTGTCAGAACTAAGAGATACCCAGGAAGTCCCTTC	60	
Db	580	GATGCCACTGTCAATCTCTGCTGATTTTGTCAAAGTAAGAGATACCCAGAAGGCCCTTC	639	
QY	61	AGGGAAACGTAAGCGTGTGTAACAACCAACCAGAAAAAACCTTCTCCTGCCAACTGTGTGA	120	
Db	640	AGGGAAACGTAAGCGTGTGTAAGGCAACCAACCAGAAAACCTTTCTTGCCAACGTGTGA	699	
QY	121	CAAGSCCTTTAACACGTGTTTGAAATAATTAAGGTTCACTTCTCTCATCAGAGAGAG	180	
Db	700	CAAGSCCTTTAACAGTGTGGAATAATTAAGGTTCACTTCTCTCATCAGCGGAGAG	759	
QY	181	GCCTTACAAGTGCATACAACAAGACTGCACCAAGSCCTTTGTTCTTAGTACAAATTACA	240	
Db	760	GCCTTACAAGTGCACACACCAAGACTGCACCAAGSCCTTTGTTCTTAGTACAAATTACA	819	
QY	241	AAGGCACATGGCTACTCATCTCTCGTGAAGAAACCCCAAGTGTAAATTATTTGAGAAAT	300	
Db	820	AAGGCACATGGCTACTCATCTCTCGTGAAGAAACCCCAAGTGTAAATTACTGTGAGAAAT	879	
QY	301	GTITTCACCGGAAAGATCATCTGAAGAATCACCTTCCATACACAGCACCTTAAACAAGAGAC	360	
Db	880	GTITTCACCGGAAGACCATCTGAAGAATCACCTTCCATACACATGACCCCAACAAGAGAC	939	
QY	361	GTITTAAGTCGAAGAAATGTGGCAAGACTACANTCAAGCTTGGATTTTAAAGCTCACATT	420	
Db	940	CTTTAAATCGAAGAGTGTGGCAAGACTACAATCCAAAGCTTGGTTTTAAGCGACATT	999	
QY	421	GGCCTTGATCGCGCAACRAAGTGTGACCTTCACTGTAAAGTATGTTTGCAAACTTTTGA	480	
Db	1000	GGCCTTGATCGCTGCGCACCAAGTGTGACCTTCACTGCAAGTGTGTTTGCAAACTTTGA	1059	
QY	481	AAGCACGGGAGTGTCTTCGGAGCACCTTAAATCTCATGCAAGCAAGTCGTCTCGTGGGT	540	
Db	1060	GAGCACAGGTGTGCTCTCTAGAGCACCTTGAATCTCACGCGGGAAGTCATCTCGGGGTGT	1119	
QY	541	TAAAGAAAAAAGCACAGTGCBAACATTTGTATCGCCGGTTCTTACACCCGGAAGGATGT	600	
Db	1120	GAAGCAAAAAAGCACCAATGTGAGCACTGCGAAGCAGGTTCTPACCCCGGAAGATGT	1179	
QY	601	CCGAGACACATGTGTGTGCACACTGGAAGAAAGGACTTCTCTGTGAGTATGTGCACA	660	
Db	1180	CCGAGACACATGTGTGTGCACACAGGAAGAAAGGACTTCTCTGTGAGTATGTGCACA	1239	
QY	661	GAGATTTTGGCGAAAGGATCACCTGACTTCGACATATGAAGAGAGTCAATCAAGAGCT	720	

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 142102)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J.Y., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,M., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 142102)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 142102)

JOURNAL
REFERENCE
AUTHORS

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 23, 2002 this sequence version replaced gi:21699524.
All repeats were identified using RepeatMasker:

JOURNAL
REFERENCE
AUTHORS

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 23, 2002 this sequence version replaced gi:21699524.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24490
Center clone name: 140_I_16

FEATURES

source	Location/Qualifiers
repeat_region	1. 142102 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="8" /map="8" /clone="RP11-140I16" /clone_lib="RP11-140I16" complement(416..699) /rpt_family="AluDb" 1097..1215 /rpt_family=" (TCTA) n" complement(1495..3567) /rpt_family="L1PA10" complement(3983..4434) /rpt_family="MLT2B1" complement(4435..5231) /rpt_family="ERVU-B1" complement(5235..5457) /rpt_family="ERVU" 5464..7077 /rpt_family="MERS2A" complement(7079..7553) /rpt_family="ERVU" complement(7554..7707) /rpt_family="L1MC4a" complement(7716..7912) /rpt_family="AluSg/x" complement(7916..8284) /rpt_family="L1MC4a" complement(8362..8565) /rpt_family="HAU1" 8566..8883 /rpt_family="AluSc" 8954..9729 /rpt_family="L1MC4a" complement(9745..15888) /rpt_family="L1PA5" 15890..16239 /rpt_family="L1MC4a" 16299..16797 /rpt_family="L1MC/D" 16798..16831 /rpt_family=" (CA) n" 16836..17389 /rpt_family="L1MC/D" 17423..17444 /rpt_family=" (A) n" 17474..18130 /rpt_family="L1MC/D" 18138..18182 /rpt_family=" (CA) n" 18341..18565 /rpt_family="L1ME" 18581..18889 /rpt_family="AluSx" 18946..19102 /rpt_family="L1ME" 19371..19681 /rpt_family="AluSx" 21480..21528

		Query Match	84.0%; Score 1260.4; DB 9; Length 142102;
		Best Local Similarity 99.9%; Pred. No. 0;	
		Matches 1261; Conservative	1; Indels 0; Gaps 0;
repeat_region	/rpt_family="G-rich"		
repeat_region	22177..22199		
repeat_region	/rpt_family="AT-rich"		
repeat_region	26803..26828		
repeat_region	/rpt_family="(T)n"		
repeat_region	27273..27305		
repeat_region	/rpt_family="(CA)n"		
repeat_region	27755..27794		
repeat_region	/rpt_family="AT-rich"		
repeat_region	27902..27929		
repeat_region	/rpt_family="(TG)n"		
repeat_region	complement(33140..33435)		
repeat_region	/rpt_family="AluXo"		
repeat_region	35118..35145		
repeat_region	/rpt_family="AT-rich"		
repeat_region	complement(38288..38621)		
repeat_region	/rpt_family="MER2"		
repeat_region	complement(39628..39944)		
repeat_region	/rpt_family="AluSx"		
repeat_region	41172..41289		
repeat_region	/rpt_family="L2"		
repeat_region	42615..42641		
repeat_region	/rpt_family="(A)n"		
repeat_region	complement(44033..44353)		
repeat_region	/rpt_family="L1PA7"		
repeat_region	44354..45325		
repeat_region	/rpt_family="L1PA7"		
repeat_region	46756..47032		
repeat_region	/rpt_family="L1ME4A"		
repeat_region	47322..48209		
repeat_region	/rpt_family="L1ME1"		
repeat_region	48210..48557		
repeat_region	/rpt_family="AluSx"		
repeat_region	48558..48920		
repeat_region	/rpt_family="L1ME1"		
repeat_region	48848..48986		
repeat_region	/rpt_family="L1MA9"		
repeat_region	complement(49958..50264)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(50322..50607)		
repeat_region	/rpt_family="AluSq"		
repeat_region	50969..51157		
Query Match	84.0%; Score 1260.4; DB 9; Length 142102;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1261; Conservative	1; Indels 0; Gaps 0;		
QY	240 AAGGCACATGGCTACTCATCTCTCTGAGAAACCCACCAAGTGTAAATTGTGAGAAA	299	
DB	31099 AAGGCACATGGCTACTCATCTCTCTGAGAAACCCACCAAGTGTAAATTGTGAGAAA	31040	
QY	300 TGTTTCACGGAAAGATCATCTGAGAAATCCTCATACACAGCCCTAACCAAGAGA	359	
DB	31039 TGTTTCACGGAAAGATCATCTGAGAAATCCTCATACACAGCCCTAACCAAGAGA	30980	
QY	360 CGTTTAAATGCGAAGATGTGGCAAACTACAAATACCAAGCTTCGATTTAAACGTCACT	419	
DB	30979 CGTTTAAATGCGAAGATGTGGCAAACTACAAATACCAAGCTTCGATTTAAACGTCACT	30920	
QY	420 TGGCCTTGATCGCGCAACAAAGTGTGACCTCCTCCTGTAAGTATGTTGCAAACTTTTG	479	
DB	30919 TGGCCTTGATCGCGCAACAAAGTGTGACCTCCTCCTGTAAGTATGTTGCAAACTTTTG	30860	
QY	480 AAACACGGAGTGCTTCTTGAGACACTTAAATCTCATCGCAAGTCTGCTGCTGGGG	539	
DB	30859 AAACACGGAGTGCTTCTTGAGACACTTAAATCTCATCGCAAGTCTGCTGCTGGGG	30800	
QY	540 TTAAGAAAAAAGACACAGTGGCAACTTGTGATCGCGGTTTCTACCCGAAAGGATG	599	
DB	30799 TTAAGAAAAAAGACACAGTGGCAACTTGTGATCGCGGTTTCTACCCGAAAGGATG	30740	
QY	600 TCCGGAGACATGGTGGTGACACATGGAAAGAGCACTTCTCTGTCAATTTGAC	659	

QY	540	TTAAAGAAAAAAGCACCAGTGGCGAACATTGTGATCGCGGTTCTACACCCGAAAGGATG	599
Db	144485	TGAAGAGAAAAAAGCACCAGTGTGAGCACTGGAGCGGAGGTTCTACACCCGGAGGATG	144544
QY	600	TCCGGAGACACATGTTGGTGGCACCTGGAAGAAAGGACTTCCTCTGTGAGTATTGTGCAC	659
Db	144545	TTCCGGAGACACATGTTGGTGGCACACGGGAAGAAAGGACTTCCTCTGTGAGTATTGTGCAC	144604
QY	660	AGAGATTTCGGGGAAGAGTACACTGATCTGACATATGAGGAAGAGTACCAATCAAGAGC	719
Db	144605	AGAGATTTCGACGAAGAGTACCTCACTCGACATATGAGGAAGAGTACCAATCAAGAGC	144664
QY	720	TTCTGAAGTCAAAAACAGAACAGTGGATTTCCTTGACCATTTTACCTGCAATGTGTCTG	779
Db	144665	TTCTGAAGTCAAAAACAGAACAGTGGATTTCCTTGACCATTTTACCTGCAATGTGTCTG	144724
QY	780	TGCTCTAAAGACGAGCTCCTTCGGGTGATGTCCTTAATCTTCCAGTGAACCTGTTATCAA	839
Db	144725	TGCTCTAAAGATGAGTCTCTCGGTGATGTCCTTTCAGCGAATCTTGTCTCAA	144784
QY	840	AGCCATTCAAAACACATTTCGAGTTAAACCTCTACAACTCCATTTTCAGTCCATGCAGA	899
Db	144785	AGCCATTCAAAACACATTTCGAGTTAAACCTCTATTAACCTCCATTTTCAGTCCATGCAGA	144844
QY	900	GCTCGGATCTGCCACCAATGATCAAACTTTACCTTTGGAAATGACATGCGCCCAATAG	959
Db	144845	GCTCTGGGTCTACTTCACCAATGATCAAACTTTACCTTTGGAAATGACATGCGCCCAATAG	144904
QY	960	ATATGGACACATGTTTATCCCTCTCACACACCTTTCTTCAATATCCGTTTCAGTCTTACCT	1019
Db	144905	ATATGGACACATGTTTATCCCTCTCACACACCTTTCTTCAATATCCGTTTCAGTCTTACCT	144964
QY	1020	CATATGCAATTTCTATTCTCTGAAAGAAAGACGCAATTAAAGGGGGAAATTGAGAGTTACC	1079
Db	144965	CATAGCAATCTCTATTCCCGAGAAAGAACAGCCCTTAAAGGGGGAAATCGAGAGTTATC	145024
QY	1080	TCATCGAGTTACAAGGTGGCGTCCCTCTTATCCCAAGATTCTCAAGCAATCTCATCGT	1139
Db	145025	TCATCGAGTTACAAGGTGGCGCACCATTCTTATCCAGCAATCTCATCGT	145081
QY	1140	CTAAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGAGACCTCTCCC	1199
Db	145082	CTAAGTTAGGGTTAGAGCCTCAGAGTGGCTCCCGACGATGGTCTGGGACCTCTCCC	145141
QY	1200	TATCCAAAAGCTTATCTCANTCAGTGACCCCTTAAACACACACGACATTTGGATTTCCTC	1259
Db	145142	TGTCAAAGAGCTCTATCTCTATCAGTGACCCCTGNAACACACGACATTTGGATTTCCTC	145201
QY	1260	AGTTGTTTAAATTTTATACACCTTTAAATGGTCTCCCTTAATCCTCTATCAGTGGGAGCC	1319
Db	145202	AGTTGTTTCAATTTTATACCAATTAATGGTCTCTTCAATCGGCTCTCCGTTGGGAAGCC	145261
QY	1320	TTGGAATGAGCTATTCGAGGAAGACACATTTCTGTTTCCGAGCTCCCAACACAAA	1379
Db	145262	TTGGCATGAGCTATTCGAGGAAGACGAGCGCATTTCTTCTGTTTCTAGCTGCCACACAAA	145321
QY	1380	CACAGGATCTTTGAGATCTCTGCAAAACACTATAGGGCTTGGGTCTCTGCACTCATGTGAG	1439
Db	145322	CACAGGATCTTTGAGATCTCTGCGNACACAGTGGGCTTGGCTCTCTGCACTCATGTGAG	145381
QY	1440	CAGCTTTTCCACGAGTTTAAAGCAAGATACCACTCCCAAGTTTCCATCAGAGCTTTTC	1499
Db	145382	CAGCTTTTCCACGAGCTTAAAGCTCAAGCACTACCTTGGCCCCGTTTCCACAGGCTTTTC	145441
QY	1500	AG 1501	
Db	145442	AG 145443	

RESULT 9

AC097274/c

LOCUS

DEFINITION

182102 bp

linear

HTG 18-JAN-2002

AC097274

Mus musculus clone RP23-59B17, WORKING DRAFT SEQUENCE, 8 unordered

```

titles.
AC097274      GI:18201773
AC097274.5   GI:18201773
HTG; HTGS_PHASEI; HTGS_DRAFT.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 182102)
McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
Bastide,M., Kuit,K., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Santos,L.,
Shah,R.S., Spiegel,L.A., Palmer,L., Yang,C. and Zutavern,T.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 182102)
McCombie,W.R.
Direct Submission
Submitted (13-OCT-2001) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Jan 18, 2002 this sequence version replaced gi:16973738.
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: RP23-59B17
Center clone name: RP23-59B17
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 48370: contig of 48370 bp in length
* 1
* 48371 43082: gap of unknown length
* * 49083 95033: contig of 45951 bp in length
* * 95034 95745: gap of unknown length
* * 95746 129069: contig of 33324 bp in length
* * 129070 129781: gap of unknown length
* * 129782 150082: contig of 20301 bp in length
* * 150083 150793: gap of unknown length
* * 150794 162439: contig of 11646 bp in length
* * 162440 163150: gap of unknown length
* * 163151 172099: contig of 8949 bp in length
* * 172100 172810: gap of unknown length
* * 172811 178596: contig of 5786 bp in length
* * 178597 179307: gap of unknown length
* * 179308 182102: contig of 2795 bp in length.
*
* Location/Qualifiers
*     1..182102
*       /organism="Mus musculus"
*       /mol_type="genomic DNA"
*       /db_xref="taxon:10090"
*       /clone="RP23-59B17"
*
Query Match          68.7%; Score 1031.6; DB 2; Length 182102;
Best Local Similarity 89.1%; Pred. No. 6,3e-292; Indels    3; Gaps    1;
Matches 1125; Conservative
Qy 240 AAGGCATCGGTACTCATTCCTCCTGAGAAAAACCACAAGTGAATTATTGTGAGAAA 299
Db 60773 AAAGGCATCGGTACTCATTCCTCCTGAGAAACCACCAAGTGAATTACTGTGAGAAAA 60714
Qy 300 TGTTTCCGGGAAGATCATCTGAAGAATCACTCCATACACGACCTTAACAAGAGA 359

```

[illegible]

Db 60713 TGTTTACCGGAAAGACCCTCTGAAGATCACCTCCATACATGACACCCCAAGAGAGA 60654

QY 360 CGTTTAAAGTCGGAAGATGCGAAGAACTACATACCAAGCTTGGATTTAAACGTCAC 419

Db 60653 CCTTTAAATCGGAAGATGCGAAGAGCTACAATACCAAGCTTGGTTTAAAGCGACACT 60594

QY 420 TGGCCTTGATCGCGCAAGAGTGGTGAACCTACCTGTAAAGTATGTTTGCACAACTTTTG 479

Db 60593 TGGCCTTGATCGCGCAAGAGTGGTGAACCTACCTGTAAAGTATGTTTGCACAACTTTTG 60534

QY 480 AAAGACGGAGTGTCTTCTGGAGCACCTTAAATCTCATGACGAGCAAGTCTGTCTGGTGGG 539

Db 60533 AGAGCACAGTGTCTCTAGACCACTGAAATCTCAGCGGCAAGTCACTTGGGGGTG 60474

QY 540 TTAAGAAAAAAGACACGATGCGAATGATCGCCGGTCTTACACCCGAAAGGATG 599

Db 60473 TGAAGAGAAAAAGACCAATGTGACACTGCGAAGCGAGTCTTACACCCGGAAGATG 60414

QY 600 TCCGAGACACATGTGTGGTGCACACTGGAAGAAAGACTTCTCTGTCACTACTGCGAC 60354

Db 60413 TCCGAGACACATGTGTGGTGCACACTGGAAGAAAGACTTCTCTGTCACTACTGCGAC 60354

QY 660 AGAGATTTGGCGAAGAGTACCTGACTCGACATATGAAGAGAGTCAATCAATCAAGAC 719

Db 60353 AGAGATTTGGCGAAGAGTACCTGACTCGACATATGAAGAGAGTCAATCAATCAAGAC 60294

QY 720 TTCTGAAGGTCAAAACAGACAGTGGATTTCTTGAACCCATTTACCTGCAATGTGTCTG 779

Db 60293 TTCTGAAGGTCAAAACAGACAGTGGATTTCTTGAACCCATTTACCTGCAATGTGTCTG 60234

QY 780 TGCCTATAAAGACAGCTCTTCCGGTGATGCTCTTACCTTCAGTGAATGTTTATCAA 839

Db 60233 TGCCTATAAAGATGAATCTCTCGGGTCATGCTCTTACCTTCAGTGAATGTTTATCAA 60174

QY 840 AGCATTCACAAACACTTTGAGTTAAACCTCTCAACACTCCATTTCAAGTCCATGACAG 899

Db 60173 AGCATTCACAAACACTTTGAGTTAAACCTCTCAACACTCCATTTCAAGTCCATGACAG 60114

QY 900 GCTCGGATCTGCCCAACAAATGATCAACTTTACCTTTGGGAATGACATGCCCAATAG 959

Db 60113 GCTCGGATCTGCCCAACAAATGATCAACTTTACCTTTGGGAATGACATGCCCAATAG 60054

QY 960 ATATGGACACTGTTTCATCCCTCTCACCACCTTCTTTCAATATCGGTTCACTACCT 1019

Db 60053 ATATGGATGTTTACCCCTCTCATCTTGTCTTCAATGCGCATTCAGTCTACCT 59994

QY 1020 CATATGCAATTTCTATTCCTGAAAGAAACAGCCATTAAGGGGGAATTCAGAGTTACC 1079

Db 59993 CATACGCAATCTCTATTCCTGAAAGAAACAGCCATTAAGGGGGAATTCAGAGTTACC 59934

QY 1080 TGATGGATTACAGTGGGCGTCCCTTCATCCCAAGATTCCTCAAGCATCGTCATCAT 1139

Db 59933 TGATGGATCTACAGTGGGCGTCCCTTCATCCCAAGATTCCTCAAGCATCGTCATCAT 59877

QY 1140 CTAAGTAGGTTGGATCTCTCAGATGGGTCCTAGATGATGTCGAGGAGACCTCTCC 1199

Db 59876 CTAAGTAGGTTAGAGCTCTCAGATGGGTCCTAGATGATGTCGAGGAGACCTCTCC 59817

QY 1200 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACAGCATGTTGATTTTCTC 1259

Db 59816 TGTCAAAGAGCTCTATCTCTATCAGTGACCCCTCAGCACACCATGATTTCTCTC 59757

QY 1260 AGTGTGTTAATTTTACCTTTAAATGGTCCCTCTATATCTCTATCAGTGGGAGCC 1319

Db 59756 AGTGTGTTAATTTTACCTTTAAATGGTCCCTCTATATCTCTATCAGTGGGAGCC 59697

QY 1320 TTGGAATGAGCTATTCCAGGAGAGACATCTCTGTTTCCAGCTCCCAACACAAA 1379

Db 59696 TTGGAATGAGCTATTCCAGGAGAGACATCTCTGTTTCCAGCTCCCAACACAAA 59637

QY 1380 CACAGGATCTTCAGGATCTGCAAAACACTATAGGCTTGGTCTCTGCACTCACTGTGAG 1439

Db 59636 CACAGGATCTTCAGGATCTGCAAAACACTATAGGCTTGGTCTCTGCACTCACTGTGAG 59577

QY 1440 CAGCTTTCACGACGAGTTTAAAGCAAGTAGTACCACCTCCACGTTTCCATCAAGCTTTTC 1499

Db 59576 CAGCTTTCACGACGAGCTGAGCTCAAGCACTACCTTGGCCCCGTTTCCACGAGCGTTTC 59517

QY 1500 AG 1501

Db 59516 AG 59515

RESULT 10
AL807387/c
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-24J10 on chromosome 4, complete
sequence.
ACCESSION
AL807387
VERSION
AL807387.10 GI:30350042
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Harrison.E.
Direct Submission
Submitted (03-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 3, 2003 this sequence version replaced gi:26788130.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-24J10 is
from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6.

FEATURES

Location/Qualifiers
1..241148
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-24J10"
/clone_lib="RPCI-23"

ORIGIN

Query Match

68.7%; Score 1031.6; DB 10; Length 241148;

Qy	240	AAAGGACATGGCTACTCACTCTCTGAGAAACCCCAAGTGAATATTTGTCAGAAA	299	75147
Db	75206	AAAGGACATGGCTACTCACTCTCTGAGAAACCCCAAGTGAATATTTGTCAGAAA		
Qy	300	TGTTTCCCGGAAAGATCATCTGAAGATCACTCCATACACAGCCCTAACAAAGAGA	359	
Db	75146	TGTTTCCCGGAAAGATCATCTGAAGATCACTCCATACACAGCCCTAACAAAGAGA		
Qy	360	CGTTTAAGTCGGAAGATGCGCAAGAACTCAATACCAAGCTTGGAATTTAAACGTCACT	419	
Db	75086	CGTTTAAGTCGGAAGATGCGCAAGAACTCAATACCAAGCTTGGAATTTAAACGTCACT		
Qy	420	TGGCCTTGATGCGCCCAAGTGGTGACCTCACTGAAGTATGTTTGCAAACTTTTG	479	
Db	75026	TGGCCTTGATGCGCCCAAGTGGTGACCTCACTGAAGTATGTTTGCAAACTTTTG		
Qy	480	AAAGCAAGGAGTGTCTTGAGAGCACTTAATCTCATGCGCAAGTGTCTGTGGGG	539	
Db	74966	AGAGCAAGGAGTGTCTTGAGAGCACTTAATCTCATGCGCGCAAGTGTCTGTGGGG		
Qy	540	TTAAAGAAAGAACACCAAGTGCAGACATGTGATCGCCGGTTCTACACCCGAAAGATG	599	
Db	74906	TGAAGGAGAAAGAACACCAATGTGACACTGCGAAGCGAGTGTCTACACCCGAAAGATG		
Qy	600	TCCGGAGACACATGGTGTGTCACACTGGAAGAAAGACTTCTCTGTGATTTGTCAC	659	
Db	74846	TCCGGAGACACATGGTGTGTCACACTGGAAGAAAGACTTCTCTGTGATTTGTCAC		
Qy	660	AGAGATTGGCGAAAGGATCACTGACTGCACATATGAGAGAGTCAAAATCAAGAGC	719	
Db	74786	AGAGATTGGCGAAAGGATCACTGACTGCACATATGAGAGAGTCAAAATCAAGAGC		
Qy	720	TTCTGAGGTCAAAACAGAACCAAGTGGATTTCTTGACCCATTTTACCTGCAATGTGTCTG	779	
Db	74726	TTCTGAGGTCAAAACAGAACCAAGTGGATTTCTTGACCCATTTTACCTGCAATGTGTCTG		
Qy	780	TGCTTATAAAGAGAGCTCTCTCCGGTGTGATGCTTACCTTCACTGAGTCACTGTATCAA	839	
Db	74666	TGCTTATAAAGAGATGAATCTCTCGGGTGTGATGCTTACCTTCACTGAGTCACTGTATCAA		
Qy	840	AGCAATTCACAAACACTTTGAGTAACTTCACTCAACACTTCCATTTTCACTGCAATGAGA	899	
Db	74606	AGCAATTCACAAACACTTTGAGTAACTTCACTCAACACTTCCATTTTCACTGCAATGAGA		
Qy	900	GCTCGGATCTGCCACCAATGATCAAACTTTTACCTTTTGGGAATGACATGCCCAATAG	959	
Db	74546	GCTCGGATCTGCCACCAATGATCAAACTTTTACCTTTTGGGAATGACATGCCCAATAG		
Qy	960	ATATGGACATGTCTCATCCCTCTCACACACTTCTTTCAAAATATCCGTTTCACTTCACT	1019	
Db	74486	ATATGGATGTCTTCAACCTCTCATCATCTTGTCTTTCAAAATATCCGTTTCACTTCACT		
Qy	1020	CATATGCAATTTCTATCTCTGAAAGAACAGCCATTAAGGGGAAATGAGAGTTTACC	1079	
Db	74426	CATATGCAATTTCTATCTCTGAAAGAACAGCCATTAAGGGGAAATGAGAGTTTACC		
Qy	1080	TGATGGAGTTACAAGTGGCGCTCTTTCATCCCAAGATTTCTCAAGCATCGTCATCAT	1139	
Db	74310	TGATGGAGTTACAAGTGGCGCTCTTTCATCCCAAGATTTCTCAAGCATCGTCATCAT		
Qy	1140	CTAAGCTAGGTTGATCTCTAGATGGTCTCTAGATGATGGTGCAGAGACCTCTCC	1199	
Db	74309	CTAAGCTAGGTTGATCTCTAGATGGTCTCTAGATGATGGTGCAGAGACCTCTCTCC		
Qy	1200	TATCCAAAGCTTATCTCCATCAGTGCACCTTAAACACACAGCATTTGGATTTTCTC	1259	
Db	74249	TGTCAAAGAGCTTATCTCTATCAGTGCACCTTAAACACACAGCATTTGGATTTTCTC		
Qy	1260	AGTTGTTTAAATTTTATACACTTTTAAATGTCCTCTCTCTTATATCTTATCAGTGGAGCC	1319	

Db	74189	AGTTGTTTCAATTTTCAATACCATTTGAAATGGTCCCTTTATAACCCACTTTTCTGATGGAGGCC	74130
Qy	1320	TTGGAATGAGCTATTTCCAGGAGAGACATCTTCTGTTTCCAGCTCCCAACAAA	1379
Db	74129	TTGGGATGAGTATTTCCAGGAGAGACATCTTCTGTTTCCAGCTCCCAACAAA	
Qy	1380	CACAGGATCTTCAGGATCTTCGAAACACTATPAGGGCTTGGGTCTCTGACATCACTGTCTG	1439
Db	74069	CACAGGATCTTCAGGATCTTCGAAACACTATPAGGGCTTGGGTCTCTGACATCACTGTCTG	
Qy	1440	CAGCTTTTCCAGCAGAGTTTAAAGCAAGTACACCTCCAGCTTCCATCAAGCTTTTC	1499
Db	74009	CAGCTTTTCCAGCAGAGCTTAAAGCAAGTACACCTCCAGCTTCCATCAAGCTTTTC	
Qy	1500	AG 1501	
Db	73949	AG 73948	
RESULT 11			
AC123210/c			
LOCUS			
DEFINITION			
Rattus norvegicus clone CH230-208111, *** SEQUENCING IN PROGRESS			
*** 5 unordered pieces.			
AC123210			
AC123210.3 GI:23101071			
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.			
Rattus norvegicus (Norway rat)			
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (bases 1 to 244955)			
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,			
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,			
Ayalalbechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,			
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,			
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,			
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,			
Cardenas, V., Carter, K., Cavazos, J., Ceasar, H., Center, A.,			
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,			
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,			
Devila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,			
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,			
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,			
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,			
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,			
Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,			
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,			
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,			
Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,			
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,			
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,			
Jacks, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,			
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,			
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,			
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,			
Lorensuhshe, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,			
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,			
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,			
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,			
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,			
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L.,			
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,			
Nwankweli, O., Okwuonu, G., Olampunsgoon, A., Pal, S., Parks, K.,			
Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.-L.,			
Plopper, F., Poindexter, A., Popovic, D., Reeves, K., Regier, M.A., Reigh, R.,			
Puro, M., Quirio, J., Rachlin, E., Reeves, K., Richards, S., Riggs, F.,			
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Rose, M., Ruiz, S.J.,			
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Shatsman, S.,			
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,			
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, D.,			
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,			

Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmami, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

REFERENCE AUTHORS TITLE JOURNAL

COMMENT

2 (bases 1 to 244955)

Worley K.C.

Direct Submission

Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 244955)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRKE

Center clone name: CH230-208L11

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 214612 bases at least Q40

Consensus quality: 217681 bases at least Q30

Consensus quality: 219976 bases at least Q20

Estimated insert size: 240534; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 135745: contig of 135745 bp in length
* 135746 135845: gap of unknown length
* 135846 239680: contig of 103835 bp in length
* 239681 239780: gap of unknown length
* 239781 240825: contig of 1045 bp in length
* 240826 240925: gap of unknown length
* 240926 242465: contig of 1540 bp in length
* 242466 242565: gap of unknown length
* 242566 244955: contig of 2390 bp in length.

Location/Qualifiers

1. .244955

FEATURES source

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-208L11"

ORIGIN
Query Match      64.2%; Score 964.2; DB 2; Length 244955;
Best Local Similarity 88.9%; Pred. No. 4.3e-272;
Matches 1054; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 317 CATCTGAAGAATCACCTCCATACACAGCACCCTAACAAAGCTTGAATTTAAACGTCCTTGGCCCTTGATGCCGCA 376
      |||
Db 223660 CATCTGAAGAATCACCTCCATACACAGCACCCTAACAAAGCTTGAATTTAAACGTCCTTGGCCCTTGATGCCGCA 376

QY 377 TGTGGCAAGAACTACAATACCAAGCTTGAATTTAAACGTCCTTGGCCCTTGATGCCGCA 436
      |||
Db 223600 TGTGGCAAGAACTACAATACCAAGCTTGAATTTAAACGTCCTTGGCCCTTGATGCCGCA 436

QY 437 ACAAGTGGTGACCTCACCTGTAAGGTATGTTTGGAAAATTTTGAAGCAACGGAGTGTCT 496
      |||
Db 223540 ACAAGTGGTGACCTCACCTGTAAGGTATGTTTGGAAAATTTTGAAGCAACGGAGTGTCT 496

QY 497 CTGAGACACCTTAAATCTCATGSCAGGCAAGTCCTCTGCGGAGGATGTCGCGAGACACATGTTG 223481
      |||
Db 223480 CTGAGACACCTTAAATCTCATGSCAGGCAAGTCCTCTGCGGAGGATGTCGCGAGACACATGTTG 223481

QY 557 CAGTGCAGAACTGATGTCGCGGTTCTACCCCGAAGAGATGTCGCGAGACACATGTTG 616
      |||
Db 223420 CAGTGCAGAACTGATGTCGCGGTTCTACCCCGAAGAGATGTCGCGAGACACATGTTG 616

QY 617 GTGCACACTGGAAGAAAGGACTTCTCTGTCAGTATTGTGCACAGAGATTTGGGCGAAG 676
      |||
Db 223360 GTGCACACTGGAAGAAAGGACTTCTCTGTCAGTATTGTGCACAGAGATTTGGGCGAAG 676

QY 677 GATCACCTGATCGACATATGAAGAAGTCAACATCAAGAGCTCTGAAGGTCAAAACA 736
      |||
Db 223300 GATCACCTGATCGACATATGAAGAAGTCAACATCAAGAGCTCTGAAGGTCAAAACA 736

QY 737 GAACAGTGAATTCCTTGACCCCAATTTACCTGCAATGTCCTGTCCTTATAAAGACGAG 796
      |||
Db 223240 GAACAGTGAATTCCTTGACCCCAATTTACCTGCAATGTCCTGTCCTTATAAAGACGAG 796

QY 797 CTCCTTCGCGTGATGTCCTTACCTTCAGTGAATGTTATCAAAAGCAATTCACAAACACT 856
      |||
Db 223180 CTCCTTCGCGTGATGTCCTTACCTTCAGTGAATGTTATCAAAAGCAATTCACAAACACT 856

QY 857 TTGAGTTAAACCTCTACAACTCCATTTTCAGTCCATGAGAGCTCGGATCTGCCAC 916
      |||
Db 223120 TTGAGTTAAACCTCTACAACTCCATTTTCAGTCCATGAGAGCTCGGATCTGCCAC 916

QY 917 CAATGATCAAACTTTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCAT 976
      |||
Db 223060 CAATGATCAAACTTTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCAT 976

QY 977 CCTCTCACCACTTTCTTCAAAATCCGTCAGTTCACCTCATATGCAATTTCTATT 1036
      |||
Db 223000 CCTCTCACCACTTTCTTCAAAATCCGTCAGTTCACCTCATATGCAATTTCTATT 1036

QY 1037 CCTGAAAAAGACAGCACTTTAAAGGGGAAATTTGAGATTTACCTGATGAGTTACAGGT 1096
      |||
Db 222940 CCTGAAAAAGACAGCACTTTAAAGGGGAAATTTGAGATTTACCTGATGAGTTACAGGT 1096

QY 1097 GCGGTGCTCTTCATCCCAAGATTTCTCAAGCATGTCATCATCTAAGCTAGGTTGGAT 1156
      |||
Db 222880 GCGGTGCTCTTCATCCCAAGATTTCTCAAGCATGTCATCATCTAAGCTAGGTTGGAT 1156

QY 1157 CCTCAGATTGGTCCCTAGATGATGGTGAGAGACCTCTCCCTATCCAAAAGCTCTATC 1216
      |||
Db 222823 CCTCAGATTGGTCCCTAGATGATGGTGAGAGACCTCTCCCTATCCAAAAGCTCTATC 1216

QY 1217 TCATCAGTGACCCCTTAAACACACAGCAGATTTTCTCAGTTGTTTAAATTCATA 1276
      |||
Db 222763 TCATCAGTGACCCCTTAAACACACAGCAGATTTTCTCAGTTGTTTAAATTCATA 1276

```

Qy	1277	CCTTTAAATGCTCCTCCTATATACCTCTATCAGTGGGAGCCTTGAATGAGCTATTCC	1336
Db	222703	CCATTAAATGCTCCTCCTTACAAATCCGCTCCCGTGGGAAGCCTTGGATGAGCTATTCC	222644
Qy	1337	CAGGAGAGACACATCTCTCTGTTTCCAGTCCCAACACAAACACAGATCTTCAGGAT	1396
Db	222643	CAGGACGAGGCGCATCTCTCTGTTTCTCAGTGGCCCAACACAAACACAGATCTTCAGGAT	222584
Qy	1397	CCTGCAACACATATAGGCTTGGGTCTCTGCACTCACTGTGAGCAGCTTTACGAGCAGT	1456
Db	222583	CCTGCAACACATATAGGCTTGGGTCTCTGCACTCACTGTGAGCAGCTTTACGAGCAGC	222524
Qy	1457	TTAAGCAAGATGACACCTCCAGCTTCCATCAAGCTTTTCAG	1501
Db	222523	CTAAGCTCAAGCACTACCTCCGCTTCCACCAAGGCTTTTCAG	222479
RESULT 12			
BC023655		1820 bp	linear
LOCUS			PRI 30-JUN-2004
DEFINITION			Homo sapiens pleiomorphic adenoma gene-like 2, mRNA (cDNA clone
ACCESSION			MGC:29597 IMAGE:4764127), complete cds.
VERSION			BC023655
KEYWORDS			MGC.
SOURCE			BC023655.2 GI:33988277
ORGANISM			Homo sapiens (human)
REFERENCE			
AUTHORS			Homo sapiens
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
			1 (bases 1 to 1820)
			Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
			Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
			Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
			Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
			Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
			Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
			Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshyayuki,S.,
			Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
			Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
			McKernan,K.J., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
			Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
			Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
			Fahney,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
			Sanchez,A., Whiting,M., Madan,A., Young,A.C., Green,E.D.,
			Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Schmutz,J., Myers,R.M.,
			Dickson,M.C., Rodriguez,A.C., Grimwood,J., Skalska,U., Smailus,D.E.,
			Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
			Schnerch,A., Schein,J.B., Jones,S.J. and Marra,M.A.
			human and mouse cDNA sequences
			Generation and initial analysis of more than 15,000 full-length
			human and mouse cDNA sequences
			Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
			12477932
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			

```
QY 458 AAGGTATGTTTGAACACTTTTGAAGCAGCGGAGTGTCTTCTGAGCACCTTAAATCTCAT 517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 678 AAGGTGTGCTGCAGACCTTTGAGAGTACCAGGCCCTCTAGAGCACCTGAAGGCCAC 737
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 GCAGCAAGTCTGCTGGTGGGTTAAGAAAAAAGCAGCAGTGGGAAATTTGTATGTCG 577
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 738 TCACGCGGGTAGCAGCGGGTCCAAAGGAGAAAGACACCCCTGTGACCACTGCGACGG 797
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 578 CGGTTCTACACCCGAAAGGATGCGGAGACATGTTGGTGCACACTGGAAGAAGAC 637
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 798 CGGTTCTACTCGTAAGGATGTACGGCGGACCTAGTGTGTGCACAGCGCGTGAAGAC 857
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 TTCCTCTGTAGTATGTGCAAGAGATTTGGCGGAAAGGATCACTGACTCGACATATG 697
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 858 TTCCTGTGCCAGTACTGTGCCAGCGGTTTGGCCGTGAAGACCACTGACGCGTATGTC 917
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 698 AAGAGAGTCAAAATCAAGAGCTTCTGAAGTCAAAAAGAGACAGAGTGGATTTCTTGAC 757
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 918 AAGAGAGCCACTCGCAGAGCTGTCAAGATCAAGACAGAGCCGCTGGACATGTTAGGC 977
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 CAATTTACTCGCAATGTGTCTGTGCTATAAAGACGAGCTCCTTCCGCTGATGCTTCA 817
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 978 CTACTCAGCTGCAGCTCCACAGTCACTGTGAAGAGAGCTGAGCCCTGTGCTGCAATG 1037
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 CTTTCCAGTGA---ACTGTTATCAAGCCATTCACAAACACTTTGACGTTAAACCTCTAC 874
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1038 GCCTCTGGGAGCTAATGGGACCAAGGCCCTTCCCTGGCATGTTGCCCATGGCATGTAT 1097
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 AACACTCCATTTGAGTCCATCAGAGCTCGGAGTCTGCCA---CCAAATGATCAACT 931
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1098 GGTGCCACATCCCTACCATGCCAGACGCGGCTGCCACACTCCCTGGTGCACAAACAG 1157
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 TTACTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCATCCCTCTCACACCTT 991
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1158 CTGCCATGGTATGAGTACCTCTGGAATCCCTACCTATCTTCTCCCGAGCTCAGCTC 1217
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 992 TCITTCAAATATCGGTTGAGTTCACCTCATPATCAATTTCTATTCCTGAAAAAAGACAG 1051
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1218 CCTCCAAATATACAGCTTGGATCTACCTCATACTTCG-----CCGACAAA 1262
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1052 CCATTAAGGGGAAATTTGAGGTTACTGATGAGTATCAAGTGGCGTCCCTCTTCA 1311
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1263 TTGCCCAAGTGGAGTGGATAGTTTCTGGCGGAGCTTCTGGAAGCGCTGTCTCTCTCA 1332
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1112 TCCCAAGATTTCAAGCATCGTCACTATCAAGCTAGGTTGGATCCTCAGATTTGGTCC 1371
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1323 TCCGCTGAACCCAGCCGCTCA-----CCTCAGCCGCGGCA 1361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1172 CTAGATGATGGTGAGAGACCTCTCCCTATCAAAAGCTCTATCTCCATCAGTGACCCC 1231
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1362 GCTGGGCGCTCTAGATGAAGCACTGTTGCCAAGAGCCCGCCAACTCTCTGAGGCC 1421
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1232 CTAAACACACCAAGCATTTGAGTTTCTCAGTGTGTTAATTTTCACTTTAAATGTCCT 1291
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1422 CTCTGGCGTCTAATGTGACTTCTCCCACTACTGGGCTTTCTTCCACTCAACCTGCC 1481
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1292 CCTATAA---TCTCTATCAGTGGGAGCTTTGGAATGAGCTATTCCAGGAGAGCA 1348
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1482 CCGTGAACCCACCTCGGGCCACAGGAGGCTGTGATGGGCTACTCCAGGCTGAGGCA 1541
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1349 CATCTCTGTTTCCAGCTGCCCAACAAACAGGATCTTCAGGATCCTGCAAAACACT 1408
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1542 CAGCCCTGTTTACCACTTTGCAAGCTCAGCTCAAGATTTCCCGAGGCTGGGAGCA 1601
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1409 ATAGGCTTGGGTCTCTGACCTACTGTGAGAGCTTTTCCAGAGATTTAAGCAAGT 1468
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1602 CTGAATTTGGGCTCTGACCTCTTGGCTCTGCTTCCAGCTGCTGGCTGAGT---AGC 1658
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1469 ACCACCTCCCAAGTTTCCATCAAGCTTTTCAG 1501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1659 ACCACCTGCTTCTTCCATCAAGCATTTCCAG 1691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 13
CQ834146

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

ORIGIN

Query Match

Best Local Similarity

Matches 823; Conservative

Score 389.8; DB 6; Length 2335;

Pred. No. 5e-103;

Mismatches 542; Indels 48; Gaps 6;

QY 98 AACTTTCCTTGCACACTGTGTGACAGGCCCTTTAAAGTGTGAGAAATTAAGTTTCA 157

Db 125 AAGTGCCAAATGTGAATTTTCGGAAACACCTTTCTCAATGGGAGAGCTGAGGCTTCA 184

QY 158 TCCTACTCTCACACAGGAGAGAGGCCCTCAAGTGCATACAAAGACTGCACCAAGGCC 217

Db 185 AGCTCCGCAACACAGAGAGAGACCATATAGTGCCTCAGCTGCACTGTGGCAAGCT 244

QY 218 TTTGTTTCTAAGTACAAATTAAGAGCAGCATGGCTACTCATTTCTCTGAGAAAACCCAC 277

Db 245 TTTGCTTCCAAATACAAAGCTGTATAGGCAATGGCCACCCACTCAGCCCAAGAACCCAC 304

QY 278 AAGTGAATTTATTTGTCAGAAATTTTTCACCGGAAAGATCATCTGAAGATCACTTCCAT 337

Db 305 CAGTGTATGTACTGTATAGATGTTTTCACCGAAGGACCATCTGGGGAACCATCTGCAG 364

QY 338 ACACAGACCCCTAACAAAGAGAGCTTTAAGTGCAGAAATGTGCAAGTACTCAATACC 397

Db 365 ACCATGATCTTACAAAGAGGCCCTCCACTGCTCTGAGTGGGTAGAAATTAACAATAGC 424

QY 398 AAGCTTGGATTTAAAGCTCACTTGGCTTTGATGCGCAACAAAGTGGTCACTTCACTGT 457

Db 425 AAGCTGGGTACCGCGCCACCTGGCCATGCATGCTGCCAGAGCGGTGACCTCAGCTGC 484

QY 458 AAGGTATGTTTCAAACTTTTGAAGACAGGAGTGTCTTCTGGAGCACTTTAAATCTCAT 517

Db 485 AAGGTGTGCTGACAGACCTTTTGAGAGTACCCAGGCCCTCTAGAGACCATCTGAGGCCAC 544

QY 518 GCAGGCAAGTCTGCTGTGGGCTTTAAAGAAAAGAACACCACTGCGAATTTGTATCGC 577

Db 545 TCACCGCGGTAGCAGGCGGTGCCAAGGAGAAAGAACCCCTCTGTGACACTGCGACCGG 604

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2004, 09:58:19 ; Search time 786 Seconds
(without alignments)

Title: US-09-242-772-116_COPY_480_1980

Perfect score: 1501

Sequence: 1 gatggcactgtcattccg.....gtttccatcaagctttttcag 1501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002as:*

7: Geneseq2002bs:*

8: Geneseq2003as:*

9: Geneseq2003bs:*

10: Geneseq2003cs:*

11: Geneseq2003ds:*

12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1501	100.0	7313	2	AAV29268
2	351.6	23.4	2334	2	AAV18481
3	351.6	23.4	2828	6	ABN86525
4	351.6	23.4	3215	11	ADM02534
5	350	23.3	2815	6	ABL61932
6	350	23.3	4632	6	ABN86524
7	348.4	23.2	3991	12	ADQ24242
8	301.8	20.1	5028	6	ABN86508
9	301.8	20.1	5028	10	ABT41834
10	294.4	19.6	2790	2	AAV18480
11	294.4	19.6	3975	6	ABN86526
12	294.4	19.6	3975	12	AD126094
13	277.6	18.5	2561	2	AAV29269
14	277.6	18.5	2561	6	ABT11028
15	276	18.4	2738	4	AAZ33549
16	267	17.8	475	4	AAI10993
17	267	17.8	475	4	ABA52644
18	267	17.8	475	4	AAI32253
19	267	17.8	475	4	ABA22432
20	267	17.8	475	4	AAK26360
21	267	17.8	475	4	AAK00905

C	22	267	17.8	475	4	ABS25951	Human liv
C	23	267	17.8	475	5	AAI00914	Probe #50
C	24	267	17.8	475	6	ABS00948	Human gen
C	25	267	17.8	1083	12	ADP28687	Adp28687 Human sec
C	26	131.4	8.8	230	4	AAI20227	Probe #10
C	27	131.4	8.8	230	4	ABA65256	Human f14
C	28	131.4	8.8	230	4	AAI45427	Human f14
C	29	131.4	8.8	230	4	ABA32357	Probe #10
C	30	131.4	8.8	230	4	AAK39414	Human bon
C	31	131.4	8.8	230	4	AAK13673	Human bon
C	32	131.4	8.8	230	4	ABS39002	Human liv
C	33	131.4	8.8	230	5	AAI05931	Probe #59
C	34	131.4	8.8	230	6	ABS13501	Human gen
C	35	126.8	8.4	324	6	ABL77615	Human ova
C	36	84.8	5.6	445	9	ACH41060	Human foe
C	37	72	4.8	72	4	AAI31082	DNA encod
C	38	72	4.8	72	8	ACC41633	Human zin
C	39	65.8	4.4	2540	10	ADC37636	Human nuc
C	40	65.8	4.4	2864	4	AAH14677	Human cDN
C	41	64.2	4.3	1449	12	ADO00620	Novel hum
C	42	64.2	4.3	1449	12	ADN99051	Novel hum
C	43	64	4.3	580	12	ACH92023	Human gen
C	44	63.4	4.2	1995	12	ADN98725	Novel hum
C	45	63.4	4.2	1995	12	ADO00294	Novel hum

ALIGNMENTS

RESULT 1

AAV29268

ID AAV29268 standard; cDNA; 7313 BP.

XX AC AAV29268;

XX DT 21-AUG-1998 (first entry)

XX DE Nucleotide sequence of human PLAG1.

XX KW Human PLAG1 gene; PLAG1; tumorigenesis gene; T-gene; PLAG2; CTNNB1;

XX KW antibody; benign tumour; malignant tumour; lymphoma; cancer;

XX KW inhibition; ss.

XX OS Homo sapiens.

XX FH Key

XX CDS

XX Location/Qualifiers

XX 481..1983

XX /*tag=

XX /product= "PLAG1 protein"

XX /transl_except= (pos: 1603..1605, aa: Gly)

XX /transl_except= (pos: 1861..1863, aa: Gly)

XX EP825198-A1.

XX 25-FEB-1998.

XX 17-JAN-1997; 97EP-00200130.

XX 22-AUG-1996; 96EP-00202339.

XX (KULB-) KU LEUVEN RES & DEV.

XX (UYGO-) UNIV GOETEBORGS HOLDINGBOLAGET AB.

XX Van De Ven WJM, Stenman KGD, Kas KP, Voz ML;

XX WPI; 1998-132252/13.

XX P-PSDB; AAW37948.

XX New tumorigenesis T-genes and proteins - useful for, e.g. preparing

XX antibodies for clinically diagnosing cells having non-physiological

XX proliferative capacity such as lipoblastomas.

XX Claim 2; Fig 4; 7lpp; English.

Qy	781	GCCTATAAAGACGAGCTCTCTCCGGTGA	1319
Db	1260	GCCTATAAAGACGAGCTCTCTCCGGTGA	
Qy	841	GCCATTCAAAACACATTTGCAGTTAAAC	900
Db	1320	GCCATTCAAAACACATTTGCAGTTAAAC	1379
Qy	901	CTCGGATCTGCCACCAAAATGATCAAA	960
Db	1380	CTCGGATCTGCCACCAAAATGATCAAA	1439
Qy	961	TATGGACACTGTTTCATCCCTCTCACC	1020
Db	1440	TATGGACACTGTTTCATCCCTCTCACC	1498

QY	1021	ATATGCAATTTCTATTTCTGAAAGAAGACAGCCAAATTAAGGGGGAATTTGAGAGTTACCT	1555
Db	1500	ATATGCAATTTCTATTTCTGAAAGAAGACAGCCAAATTAAGGGGGAATTTGAGAGTTACCT	1555
QY	1081	GATGGAGTTACAAGGTGGCGTCCCTCTTTTCATCCCAAGATTTCTCAAGCATCGTCATCATC	1144
Db	1560	GATGGAGTTACAAGGTGGCGTCCCTCTTTTCATCCCAAGATTTCTCAAGCATCGTCATCATC	1611
QY	1141	TAAAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGAGGAGACCTCTCCCT	1201
Db	1620	TAAAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGAGGAGACCTCTCCCT	1671
QY	1201	ATCCAAAAGCTCTATCTCCATCAGTAGCCCTCTAAACACACACAGCATTTGGATTTTCTCA	1266
Db	1680	ATCCAAAAGCTCTATCTCCATCAGTAGCCCTCTAAACACACACAGCATTTGGATTTTCTCA	1731
QY	1261	GTGTGTTAATTTCAATACCTTTAAATGGTCCCTCCCTAATATCTCTATCAGTGGGAGCCT	1332
Db	1740	GTGTGTTAATTTCAATACCTTTAAATGGTCCCTCCCTAATATCTCTATCAGTGGGAGCCT	1397
QY	1321	TGGAATGAGCTATTCACAGGAGAGACATCTTCTGTCTTCCAGCTCCCAACAAC	1388
Db	1800	TGGAATGAGCTATTCACAGGAGAGACATCTTCTGTCTTCCAGCTCCCAACAAC	1855
QY	1381	ACAGGATCTTCAGGATCTTCGAAACACTATAGGGCTTGGGTCTCTGCACCTCAGTGTGAGC	1444
Db	1860	ACAGGATCTTCAGGATCTTCGAAACACTATAGGGCTTGGGTCTCTGCACCTCAGTGTGAGC	1911
QY	1441	AGCTTTACACAGCAGTTTAAGCAAGATACCCCTCCCAAGTTTCCATCAAGCTTTTCA	1501
Db	1920	AGCTTTACACAGCAGTTTAAGCAAGATACCCCTCCCAAGTTTCCATCAAGCTTTTCA	1919
QY	1501	G 1501	
Db	1980	G 1980	
RESULT 2			
AAV18481			
ID	AAV18481 standard; cDNA; 2334 BP.		
XX	AAV18481;		
XX	18-AUG-1998 (first entry)		
DT	Human zinc finger protein (hzac).		
DE			
XX	Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;		
XX	inducible; alzheimer's disease; nuclear transcription factor; apoptosis;		
KW	cell cycle; neuronal disorder; human zinc finger; hZAC; ss.		
KW			
XX	Homo sapiens.		
OS			
XX	Key		
XX	Location/Qualifiers		
FT	803..2194		
FT	/*tag= a		

RESULT 2	
AAV18481	
ID	AAV18481 standard; cDNA; 2334 BP.
XX	
XX	AAV18481;
XX	
XX	
DT	18-AUG-1998 (first entry)
XX	
DE	Human zinc finger protein (hzac).
XX	
XX	Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
KW	inducible; alzheimer's disease; nuclear transcription factor; apoptosis;
KW	cell cycle; neuronal disorder; human zinc finger; hzac; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	803..2194
CD	/**tag= a
FT	

Qy 721 TCTGAAGTCAAAACAGAACAGTCGATTTCCTGACCCATT 762
|||
Db 1779 GATGAACAGAGCTTGCAGACCGGAGACCTTCGAGACCTT 1820

RESULT 5

ABL61932
ID ABL61932 standard; DNA; 2815 BP.

AC ABL61932;

DT 15-MAY-2002 (first entry)

Colon adenocarcinoma related gene sequence SEQ ID NO: 269.

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.

OS Homo sapiens.

AX
PN
WO200194629-A2

PD 13-DEC-2001

XX
PF
30-MAY-2007:XX
PR 05-JUN-2000: 2000US-0309473D

PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P

PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P

PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.

PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000: 2000US-0234567D
PR 22-SEP-2000: 2000US-0234567D

PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234934P

PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P

PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P

PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P

PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P

27-SEP-2000; 2000US-0235840P.
27-SEP-2000; 2000US-0235863P.

28-SEP-2000; 2000US-0236028P.
28-SEP-2000; 2000US-0236032P

28-SEP-2000; 2000US-0236033P.
28-SEP-2000; 2000US-0236034P.

28-SEP-2000; 2000US-0236109P.
28-SEP-2000; 2000US-0236111P.

29-SEP-2000; 2000US-0236842P.
29-SEP-2000; 2000US-0236891P.

02-OCT-2000; 2000US-0237172P.
02-OCT-2000; 2000US-0237173P.

02-OCT-2000; 2000US-0237278P.
02-OCT-2000; 2000US-0237294P.

02-OCT-2000; 2000US-0237295P.
02-OCT-2000; 2000US-0237316P.

03-OCT-2000; 2000US-0237425P.
03-OCT-2000; 2000US-0237598P.

03-OCT-2000; 2000US-0237604P.
03-OCT-2000; 2000US-0237606P.

03-OCT-2000; 2000US-0237608P.
01-NOV-2000; 2000US-0244867P.

01-NOV-2000; 2000US-0245084P.
X

AVALON PHARM. (HVAL-)

XX PI PI XX DR

Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,
Soppet DR, Weaver Z;
WPI; 2002-189264/24.

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.

Claim 1; SEQ ID NO 269; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABU61664 to ABU70110), or is at least 93% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, adenocarcinoma, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour.

Sequence 2815 BP; 816 A; 678 C; 572 G; 749 T; 0 U; 0 Other;

Query Match	23.3%;	Score 350;	DB 6;	Length 2815;
Best Local Similarity	69.7%;	Pred. No. 8.4e-100;		

QY	64	GAACGTTAAGCGTGTGTGAACCAAAACCAAGAAAAAACTTTCTCTGCCAACTGTGTGCACA	123
Db	196	GAAGAAGGAGGTGTGAGAAGCAAAAGCCCATGSCCAGCTTCCCTGCCAGTTATGTGGCAA	255
QY	124	GGCCCTTTAACAGTGTGTGAGAAATTAAGAGTTCACTCTCTACACAGGAGAGAGGCC	183
Db	256	GACGTTCTCTACCCCTGGAGAAAGTTCACGATTCACAATTAATCCCACTCCAGCGAGCGGCC	315
QY	184	CTACAAGTGCATACAACAAGACTGCACCAAGGCCCTTGTCTCTAAGTACAAATTAACAAG	243
Db	316	GTACAAGTGTGCGAGCCTGACTGTGGCAAGCCCTTGTGTTCCAGATATAAATTCATGAG	375
QY	244	GCATATGGCTACTCATTTCTCTGTGAGAAACCAACCAAGTGTAAATTAATGTGAAAAATGTT	303
Db	376	GCATATGGCTACCCATTCTCCCGAAGATCTCACAGTGTGCTCACTGTGAGAAGACGTT	435
QY	304	TCACCGGAAAGATCATCTGGAAGATACCTCCATACACACACCCCTAACAAAGACGTT	363
Db	436	CAACCGGAAGACCACTGTGAAAAACCACTTCCAGCCACACGCCCAACAAAATGGCCCTT	495
QY	364	TAAGTCGGAAGAAATGTGGCAAGACTACAATPACCAAGCTTGGATTTAAACGTCACCTGC	423
Db	496	TGGTGTGTGAGAGTGTGGGAAGAAGTACACACCATGCTGGCTATAAGAGGCACCTGCG	555
QY	424	CTTGCAATGCCGACAAAGTGGTGACCTCACTGTAAAGTATGTTTGCAAACTTTTGAAG	483
Db	556	CTCTCATGCCGCCAGCAGTGGGAGCCCTCACCTGTGGGTCTGTGCCCTGGAGCTAGGAG	615
QY	484	CACGGGAGTCTTCTGGAGCACCTTAAATCTCATGC---AGGCAAGTCGTTGTGTGGGT	540
Db	616	CACCGAGGTGCTACTGCAACCACTTAAGCCCATGCGGAAGAGAGGCCCTACGCGAAC	675
QY	541	TAAAGAAAAAAGCAACAGTGCGAACATTGTGATCGCCGGTTCTACACCCGGAAGGATGT	600
Db	676	CAAGGAAAAAAGCAACAGTGCACCACTGTGAAGATGCTTCTACNCCGGAGGAATGT	735
QY	601	CCGGAGACATCGTGGTGCACTGTGAAGAAAGGACTTCCTCTGTTCAGTATTGTGCACA	660

QY 281 TGTAAATTATTTGTGAGAAAATGTTTTCACCGGAAGATCATCTGAAGATCACTCCATACA 340
Db |||||
737 TGCACCTATTGTGAAAGACTTTCAACCGGAAGGATCATCTGAAGATCACTCCAGACC 796
QY 341 CACGACCTTAACAAGACAGCTTTTAAGTGCAGAAATGTGGCAAGAACTACAATACCAAG 400
Db |||||
797 CACGATCCCAACAAGATGATCTACGCTTGGAGATTTGGCAAGAAATACCAACCATG 856
QY 401 CTGGATTAAACGTCACCTTGGCTTGCATGCCGCAACAAGTGGTGACCTACCTGTAA 460
Db |||||
857 CTGGCTACAGAGGACATGATGCCCTGCATTCGGCCAGCAGCGGATCTACCTGCGGC 916
QY 461 GTATGTTGAAACTTTTGAAGCAGCGGAGTGCTTCTGGACACCTTAAATCTCATG- 519
Db |||||
917 GTCTGACCTTGGAGTGGGAGCAGCGAGTCTTGGTGCACCATCTAAGTCTCAGCG 976
QY 520 --AGCAAGTCTCTGTGGGTAAAGAAAAAAGCACCAGTGGCAACATTTGTATGC 577
Db |||||
977 GAAGAAAAGGCCACACCGCCCGAGGAGAAACACCATGCGACCACTGGAGAGA 1036
QY 578 CGGTTCTACACCCGAAGATGTCGGAGACATGCTGGTGCACACTGGAAAGGAC 637
Db |||||
1037 TGTCTTACACCCGGAAGATGTCGCTGCCACCTGGTGCACAGAGATGCAAGGAC 1096
QY 638 TTCCTCTGTCAGTATTTGTGCACAGATTTGGGCGAAAGGATCACCTGACTCGACATG 697
Db |||||
1097 TTCCTGTGTCAGTCTTGGCGCCAGAGATTTGGGCGCAAGACCACTCTCTCACACC 1156
QY 698 AAGAAGAGTCACAATCAAGAGCTTCTGAAGTCAAAACAGAACAGTGGATTTCTTGC 757
Db |||||
1157 AAGAAGACCACTCCAGAGCTGATGCAAGAGAGCTGCAAGCAGGAGATACCAAGGC 1216
QY 758 CCATTTACCTGCAATGTCT 778
Db |||||
1217 GGTACCACCACTTGGGCT 1237

RESULT 9
ABT41834
ID ABT41834 standard; DNA; 5028 BP.
XX
AC ABT41834;
XX
DT 26-JUN-2003 (first entry)
XX
DE Toxicity modelling related rat gene SEQ ID No 1536.
XX
KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW database; drug screening; toxicity assay; rat; ds.
XX
OS Rattus norvegicus.
XX
PN WO200295000-A2.
XX
PD 28-NOV-2002.
XX
PF 22-MAY-2002; 2002WO-US016173.
XX
PP 22-MAY-2001; 2001US-0292335P.
XX
PR 13-JUN-2001; 2001US-0297523P.
PR 19-JUN-2001; 2001US-0298925P.
PR 10-JUL-2001; 2001US-0303807P.
PR 10-JUL-2001; 2001US-0303808P.
PR 10-JUL-2001; 2001US-0303810P.
PR 28-AUG-2001; 2001US-0315047P.
PR 28-AUG-2001; 2001US-0324928P.
PR 22-OCT-2001; 2001US-0330462P.
PR 01-NOV-2001; 2001US-0330867P.
PR 21-NOV-2001; 2001US-0331805P.
PR 06-DEC-2001; 2001US-0336144P.
PR 19-DEC-2001; 2001US-0340873P.
PR 21-FEB-2002; 2002US-0357842P.
PR 21-FEB-2002; 2002US-0357843P.

21-FEB-2002; 2002US-0357844P.
15-MAR-2002; 2002US-0364134P.
08-APR-2002; 2002US-0370144P.
08-APR-2002; 2002US-0370206P.
08-APR-2002; 2002US-0370247P.
17-APR-2002; 2002US-0372794P.
21-APR-2002; 2002US-0371679P.
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
PI WPI; 2003-148464/14.
XX
DR Predicting at least one toxic effect of a compound, useful for toxicity
XX modelling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
PT
XX Example 4; Page; 446pp; English.
XX The invention relates to a novel method of predicting at least one toxic
XX effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the renal
CC progression of a toxic effect of a compound, identifying toxicity markers in tissues or
CC cells exposed to known renal toxin. The genes are useful as toxicity or
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
XX Sequence 5028 BP; 1289 A; 1273 C; 1058 G; 1408 T; 0 U; 0 Other;
SQ
Query Match 20.1%; Score 301.8; DB 10; Length 5028;
Best Local Similarity 66.2%; Pred. No. 2.5e-84;
Matches 451; Conservative 0; Mismatches 227; Indels 3; Gaps 1;
QY 101 TTTCTTGGCAACTGTGTGACAGGCTTTAACAGTGTGAGAAATTAAGGTTTCACATCC 160
Db |||||
557 TTTCTGTCAAAAATGCGCAAGTCTCTCTCCCTGGAGAGTTCACCATCCACAT 616
QY 161 TACTCTCACACAGAGAGAGGCCCTACAAGTGCATACAACAGACTGCACCAAGGCTTT 220
Db |||||
617 TATTTCCACACACAGGAGCGCCCATTAAGTGTCTCAAGACTGAGTGTGGCAAGGCTTC 676
QY 221 GTTCTTAAGTACAAATTAACAAGGCACATGGCTACTCTCTCTCTGAGAAACCCACAG 280
Db |||||
677 GTCTTCAAGTATAGCTGATGAGACACATGGCTACGCACCTCTCCCAAGACGCCACG 736
QY 281 TGTAAATTATTTGAGAAAAATGTTTACCGGAAGATCATCTGAAGATCACTCCATACA 340
Db |||||
737 TGCACCTATTGTGAAAGACTTTCAACCGGAAGGATCATCTGAAGATCACTCCAGACC 796
QY 341 CACGACCTTAACAAGACAGCTTTTAAGTGCAGAAATGTGGCAAGAACTACAATACCAAG 400
Db |||||
797 CACGATCCCAACAAGATGATCTACGCTTGGAGATTTGGCAAGAAATACCAACCATG 856
QY 401 CTGGATTAAACGTCACCTTGGCTTGCATGCCGCAACAAGTGGTGACCTACCTGTAA 460
Db |||||
857 CTGGCTACAGAGGACATGCTGCTGCTGCAATTCGGCCAGCAGCGGATCTACCTGCGGC 916
QY 461 GTATGTTGCAACTTTTGAAGCAGCGGAGTGCTTCTGGACACCTTAAATCTCATG- 519
Db |||||
917 GTCTGACCTTGGAGTGGGAGCAGCGAGTCTTGGTGCACCATCTAAGTCTCAGCG 976
QY 520 --AGCAAGTCTCTGTGGGTAAAGAAAAAAGCACCAGTGGCAACATTTGTATGC 577

977 GAAGAAAGGCCACACAGCGCCCGGAGGAGAGAAACACCAAGTCGACCACTGGAGAGA 1036
 578 CGGTTCTACACCCGAAAGGATGTCCGGAGACACATGGTGGTGACACTGGAAGAAAGGAC 637
 1037 TGCTTCTACACCCGGAAGGATGTGGTCGACCTGTGTCCACAGAGATGCAAGGAC 1096
 638 TTCTCTGTCTAGTATTGTGCACAGAGATTGGGGGAAAGGATCACTGACTCGACATATG 697
 1097 TTCTGTGTCTAGTATTGTGCACAGAGATTGGGGGAAAGGATCACTGACTCGACATATG 1156
 698 AAGAAGAGTCACAAATCAAGAGCTTCTGAAGGTCAAAACAGAACAGTGGATTTCCTTGAC 757
 1157 AAGAAGACCACTCCACAGAGCTGTATGCAAGAGAGCTGCAAGAGAGAGATACCAAGGC 1216
 758 CCATTACCTGCAATGTGTCT 778
 1217 GGTACCAACCCATTGGCCT 1237

RESULT 10

AAV18480

ID AAV18480 standard; cDNA to mRNA; 2790 BP.

XX AC

XX AAV18480;

XX DT 18-AUG-1998 (first entry)

XX DE BOP1 cDNA.

XX KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;

XX KW inducible; Alzheimer's disease; nuclear transcription factor; apoptosis;

XX KW cell cycle; neuronal disorder; ss.

XX OS Mus sp.

XX FH Key

XX FT CDS Location/Qualifiers

XX FT 542..2545

XX FT /*tag= a

XX FT /product= "BOP1 protein"

XX XN WO9813489-A1.

XX PD 02-APR-1998.

XX XX 22-SEP-1997; 97WO-EP005198.

XX XX 23-SEP-1996; 96US-00718661.

XX XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX XX (CNRS) CENT NAT RECH SCI.

XX PI Spengler D, Journot L;

XX DR WPI; 1998-230701/20.

XX DR P-PSDB; AAW48760.

PT New isolated tumour suppressor gene - useful for developing products for
 PT use in diagnosis and treatment of tumour(s) or neuronal disorder(s).
 XX Claim 1; Page 72-76; 118pp; English.

The present sequence represents the BOP1 cDNA isolated from the mouse
 corticotroph pituitary tumour cell line AtT-20 cDNA library. The protein
 encoded by the BOP1 cDNA displays a tumour suppressing activity when it
 was constitutively and inducibly expressed in tumour cells. The BOP1 cDNA
 and the protein it encodes are claimed to be useful in the preparation of
 therapeutic compositions, useful for treating, preventing or delaying the
 recurrence of a tumour or neuronal disorders, e.g. genetic diseases or
 acquired degenerative encephalopathies such as Alzheimer's disease. The
 BOP1 protein is also claimed to be able to induce apoptosis resulting in
 inhibition of tumour cell growth, to suppress tumour formation, to induce
 G1 arrest of the cell cycle and to act as nuclear transcription factor

XX SQ Sequence 2790 BP; 567 A; 783 C; 714 G; 626 T; 0 U; 0 Other;
 Query Match 19.6%; Score 294.4; DB 2; Length 2790;
 Best Local Similarity 64.6%; Pred. No. 4.1e-82;
 Matches 455; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
 QY 78 GTGAACCAACCAAGAAACCTTTCCTTGGCCAACTGTGTGACAAAGCCCTTTAACAGTG 137
 Db |||||
 528 GAGAAACCAAGGCCCATGGCTCCATTTCGCTGTCAAAATGTGCAAGTCTTCGTACCC 587
 QY |||||
 138 TTGAGAAATTAAGGTTTCACTCTCTACTCTACACAGAGAGAGCCCTTCAAGTGCATAC 197
 Db |||||
 588 TGAGAGATTCAACATTCACATTTATTCACATTCCTCCAGGGAGCGCCATTCAGTGTCTGA 647
 QY |||||
 198 AACAAGACTGCACCAAGGGCTTTGTTTCTAAGTACAAATTTACAAAGGCACATGGCTACTC 257
 Db |||||
 648 AGGCTGAGTGTGGCAAGCCTTCGCTCTCCAAGTATTAAGCTATGAGACACATGGCCACAC 707
 QY |||||
 258 ATTCTCTGAGAAACCCCAAGTGTAATTTGTGAGAAATGTTTCAACCGAAAGATC 317
 Db |||||
 708 ACTCGCCACAGAGATTTCACCAAGTGTACTCACTGTGAGAAGACATTCACACCGAGSACC 767
 QY |||||
 318 ATCTGAAGAAATCACCTCCATACACAGCCCTTAACAAAGAGAGAGTTTAAGTGCAGAAAT 377
 Db |||||
 768 ACCTGAAGAACCACTCCAGACCCAGATCCCAACAAAGATCTCTACGGGTGTGACGATT 827
 QY |||||
 378 GTGGCAAGAACTACAATACCAAGCTTGGATTTAACGCTCACTTGGCCTTGCATGCCGAA 437
 Db |||||
 828 GCGGCAAGAAGTACACACCACTGCTGGCTACAGAGGACCTGCGCTTGCATCGGCGA 887
 QY |||||
 438 CAAAGTGGTGACCTCACCTGTAAAGTATGTTTGCAAACTTTTGAAGACAGCGGAGTGCTTC 497
 Db |||||
 888 GCAATGGCGATCTCACCTGTGGGTGTGCACCTGGAGCTGGGAGACCCGAGGTCCTGC 947
 QY |||||
 498 TGGAGCACTTAAATCTCATGC--AGGCAAGTCTCTGTGGGGTTAAAGAAAAAAGC 554
 Db |||||
 948 TGGACCACCTCAAGTCTCACGGGAAGAAAGGCCAACCCAGCCACCCAGGAGAGAAAT 1007
 QY |||||
 555 ACCAGTGCAGAACTTGTGATCGCGGTTCTACACCCGAAAGAGTGTCCGGAGACATGG 614
 Db |||||
 1008 ACCAGTGCACCACTGTGATAGATGCTTCTACACCCGAAAGAGTGTGCTGCCACCTGG 1067
 QY |||||
 615 TGGTGCACACTGGAAGAAAGACTTCTCTCTCAGTATTTGTGCACAGAGATTGGGCGAA 674
 Db |||||
 1068 TGGTCCACACAGGATGCAAGGACTTCTCTGTCTCAGTCTGTGCCAGAGATTGGGCGCA 1127
 QY |||||
 675 AGGATCACCTGACTCGACATATGAAGAGAGTCAATCAAGACTTCTGAAGTCAAAA 734
 Db |||||
 1128 AAGACCACCTCACTCGTCCACACCAAGAGACCCCACTCCAGGAGCTGATGCAAGAGATA 1187
 QY |||||
 735 CAGAACCACTGGATTTCCTTGGACCCCAATTTACCTGCAATGTGTCT 778
 Db |||||
 1188 TGCAGGCGAGGATTACCAAGAGCAATTTCCAACTCATTTGGCCT 1231

RESULT 11

ABN86526

ID ABN86526 standard; cDNA; 3975 BP.

XX AC

XX ABN86526;

XX XX 21-OCT-2002 (first entry)

DT Nucleotide sequence of mouse ZAC zinger finger protein cDNA.
 XX Cardiovacular; Flt-1; CD44; Lot-1; AA892598; Mrg-1; cardiant; mouse;

KW cerebroprotective; antiarteriosclerotic; ZAC; zinger finger protein; ss.
 XX Mus musculus.

XX OS

XX WO200238794-A2.

XX XN

615 TGGTGACACCTGGAGAGAGGACTTCTCTCTGAGTATTGTGCACAGAGATTGGCGGAA 674
 775 TGGTCCACACAGGATGCAAGGACTTCTCTCTGAGTATTGTGCACAGAGATTGGCGGCA 834
 675 AGGATCACCTGACCTGACATATGAAGAGAGTCAATCAAGAGCTTCTGAAGTCAAAA 734
 835 AAGACCACTCTCTGTCACACCAAGAGAGCCACTCCAGGAGCTGATGCAAGAGAATA 894
 735 CAGAACCACTGAGTCTCTGACCCATTTACTGCAATGTGCT 778
 895 TGCAGGAGGAGATTACCAAGAGCAATTTCCAACTCATTGCGCT 938

RESULT 12
 ADI26094
 ID ADI26094 standard; cDNA; 3975 BP.
 XX AC ADI26094;
 XX DT 22-APR-2004 (first entry)
 XX DE Human cDNA encoding protein that promotes STAT6 activation #30.
 XX KW ss; gene; human; signal transducer and activator of transcription 6;
 XX KW STAT6; immunogen; STAT6 activation; allergy; inflammation;
 XX KW autoimmune disease; diabetes; hyperlipidemia; infection; cancer;
 XX KW Th1 hyperactive disease; rheumatoid arthritis; osteoarthritis;
 XX KW systemic lupus erythematosus; sepsis; asthma; allergic rhinitis;
 XX KW ischaemic heart disease; subarachnoid haemorrhage; viral hepatitis; AIDS.
 XX OS Homo sapiens.
 XX PN WO2003104277-A2.
 XX PD 18-DEC-2003.
 XX PF 05-JUN-2003; 2003WO-JP007123.
 XX PR 05-JUN-2002; 2002JP-00164257.
 XX PR 06-JUN-2002; 2002US-0385912P.
 XX PR 26-DEC-2002; 2002JP-00377326.
 XX PR 27-DEC-2002; 2002US-0436467P.
 XX PR 15-MAY-2003; 2003JP-00137505.
 XX PR 16-MAY-2003; 2003JP-0470836P.
 XX PA (ASAH) ASAH KASEI KK.
 XX PI Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;
 XX WI P-I; 2004-122214/12.
 XX WI P-PSDB; ADI26095.
 XX PT New signal transducer and activator of transcription 6 activation
 XX PT promoting purified protein, for diagnosing and treating disease
 XX PT associated with activation/inhibition of transcription factor e.g.
 XX PT diabetes and cancer.
 XX PS Claim 4; SEQ ID NO 59; 1368pp; English.
 XX CC The invention relates to a purified protein promoting signal transducer
 XX CC and activator of transcription 6 activation (STAT6). The protein is
 XX CC useful for the producing an antibody, which involves administering the
 XX CC protein or its epitope-bearing fragments to a non-human animal as an
 XX CC antigen. The nucleic acid is useful for diagnosing a disease or
 XX CC susceptibility to a disease related to expression or activity of the
 XX CC protein. A transformant expressing the protein is useful for screening
 XX CC compounds which inhibit or promote STAT6 activation. A transformant
 XX CC expressing the protein is useful for producing a pharmaceutical
 XX CC composition. Compositions, antibodies and antisense molecules are useful
 XX CC for the treating a disease associated with STAT6 activation such as
 XX CC allergic diseases, inflammation, autoimmune diseases, diabetes, AIDS,
 XX CC hyperlipidemia, infections disease and cancers. Compositions are useful
 XX CC for treating disease associated with STAT6 activation and/or prevention

16-MAY-2002.
 08-NOV-2001; 2001WO-US046816.
 09-NOV-2000; 2000US-0247457P.
 (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 Lee RT;
 WPI; 2002-590446/63.
 Diagnosing cardiovascular conditions e.g. stroke, myocardial infarction
 in a subject, by detecting expression of a nucleic acid molecule such as
 Fit-1, CD44, Lot-1, AA892598 and Mrg-1 in biological sample from subject.
 Disclosure; Page 108-109; 113pp; English.
 The invention relates to diagnosing a cardiovascular condition
 characterised by aberrant expression of a nucleic acid molecule (I) such
 as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 or its expression product. One
 method involves contacting a biological sample from a subject with an
 agent which specifically binds to (I), its expression product or a
 fragment of an expression product and measuring the amount of bound
 agent. The method is useful for diagnosing a cardiovascular condition
 such as myocardial infarction, stroke, arteriosclerosis, heart failure,
 and cardiac hypertrophy. Methods useful for determining the stage of
 cardiovascular condition and pharmaceutical compositions for treating the
 above cardiovascular conditions are also provided. The present sequence
 represents a mouse ZAC zinger finger protein cDNA
 Sequence 3975 BP; 1010 A; 1001 C; 880 G; 1084 T; 0 U; 0 Other;
 Query Match 19.6%; Score 294.4; DB 6; Length 3975;
 Best Local Similarity 64.6%; Pred. No. 5e-82;
 Matches 455; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
 78 GTGAACCAACCAAGAAAACCTTCTGTCGCACTGTGTGCAAGGCTTTAACAAGTG 137
 235 GAGAACCAAGGCGCATGGCTCCATTCGCTGTCGCAAGTGTGCAAGGCTTTAACAAGTG 294
 138 TTGAGAAATTAAGGTTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 197
 295 TGGAGAAAGTTACCAATTCACAAATTTATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 354
 198 AACAGACTGCAACCAAGGCTTTGTTTCTAAGTACAAATTAACAAGGCACTGGCTACTC 257
 355 AGGCTGAGTGTGGCAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 414
 258 ATTCTCTGAGAAACCAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 317
 415 ACTGCCACAGAAATTCACCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 474
 318 ATCTGAAGAACTCACT 377
 475 ACCTGAAGAACTCACT 534
 378 GTGGCAAGAACTCACT 437
 535 GCGGCAAGAACTCACT 594
 438 CAAAGTGTGCACT 497
 595 GCAATGGGCACT 654
 498 TGGAGCACTTAAATCTCTATGC---AGGCAAGTCTGTCTGTGGGTGTTAAGAAATAAAGC 554
 655 TGAACCACT 714
 555 ACCAGTGTGCACT 614
 715 ACCAGTGTGCACT 774

Db 411 AGCACGAGGTGCTACTGGACCACCTCAAGCCCATCGGAAGAGAGCCCTAGCGGA 470
Qy 539 GTTAAAGAAAAGACACAGTGGGAAATTTGGATCGCCGGTCTACACCGGAAGGAT 598
Db 471 ACCAAGGAAAAGACACAGTGGGAAATTTGGATCGCCGGTCTACACCGGAAGGAT 530
Qy 599 GTCCGGAGACATGGTGGTGCACACTCGAAGAGAGGACTTCTCTGTCAGTATTGTGCA 658
Db 531 GTGCGACGCCACCTGGTGGTCCACAGAGATCAAGGACTTCTGTGCGGATTTCTGTGCC 590
Qy 659 CAGAGATTGGCGGAAAGGATCACCTGACTCGACATATGAAGAAGAGTCAACATCAAGAG 718
Db 591 CAGAGATTGGCGGAAAGGATCACCTGACTCGACATATGAAGAAGAGTCAACATCAAGAG 650
Qy 719 CTCTGAAAGGTCAAAACAGAACAGTGGGATTTCTTGACCCATT 762
Db 651 CTGATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGCACCTT 694

RESULT 14
ABT11028
ID ABT11028 standard; cDNA; 2561 BP.
XX AC ABT11028;
XX
XX
XX 04-DEC-2002 (first entry)
XX
XX Human breast cancer associated coding sequence SEQ ID NO: 1162.
XX Human; breast specific gene; breast cancer; differential expression;
XX cytoskeletal; gene therapy; gene; ss.
XX Homo sapiens.
XX
XX WO200259271-A2.
XX
XX 01-AUG-2002.
XX
XX 25-JAN-2002; 2002WO-US002176.
XX
XX 25-JAN-2001; 2001US-0263757P.
XX 23-APR-2001; 2001US-0286090P.
XX 23-MAY-2001; 2001US-0292517P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Orr MS, Nation M, Diggins JC, Zeng W;
XX
XX WPI; 2002-674803/72.
XX
XX Diagnosing breast cancer in a patient comprises detecting the level of
XX gene expression in cell or tissue samples, where a differential gene
XX expression is indicative of breast cancer.

PS Claim 1; SEQ ID NO 1162; 260pp + Sequence Listing; English.
XX
XX The present invention relates to methods of diagnosing breast cancer in a
XX patient, which comprise detecting the level of expression in a tissue
XX sample of two or more genes selected from those shown in AB09867-
XX ABT1112, where a differential expression of the genes indicates breast
XX cancer. The methods are useful in diagnosing, treating, detecting the
XX progression, and in monitoring treatment of breast cancer in patients.
XX The methods are also useful as a screening tool for agents that modulate
XX the onset or progression of breast cancer. The breast cancer genes may be
XX used as diagnostic markers for the prediction or identification of the
XX malignant state of breast tissue, for confirming the type and progression
XX of cancer, and for drug screening and assays. The present sequence is a
XX coding sequence of the invention. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub.published_pct_sequences
XX
XX Sequence 2561 BP; 704 A; 624 C; 516 G; 717 T; 0 U; 0 Other;

Query Match 18.5%; Score 277.6; DB 6; Length 2561;
Best Local Similarity 71.9%; Pred. No. 8.7e-77;
Matches 377; Conservative 0; Mismatches 144; Indels 3; Gaps 1;
Qy 242 AGGCACATGGCTACTCATTTCTCTGAGAAAAACCCCAAGTGTAAATTTGTGAGAAATG 301
Db 171 AGGCATATGGCTACCATTTCTCCAGAAATCTCCACAGTGTGCTCACTGTGAGAAAGG 230
Qy 302 TTTCACCGGAAAGATCATCTGAGAAATCACCTCCATACACAGCCCTAAACAAAGAGAG 361
Db 231 TTCAACCGGAAAGACCATCTGAAAAACCACTCCAGACCCAGACCCCAAAATGGCC 290
Qy 362 TTTAAGTGGGAAGATGTGGCAAGAACTACAACTACCAAGCTTGGATTTAAACGTCACTTG 421
Db 291 TTTGGGTGTGAGGAGTGTGGGAAGAGTACAAACCATGCTGGCTATTAAGAGGCACCTG 350
Qy 422 GCCTTGCATGCCGCAACAAAGTGTGACCTCACCTGTAAAGTATGTTTGCATACTTTGAA 481
Db 351 GCCCTCCATGCCGCCAGCAGTGGGACCTCACCTGTGGGTCTGTGCCCTGAGCTAGGG 410
Qy 482 AGCACGGGAGTGTCTTCTGAGACACCTTAAATCTCATGC--AGCAAGTCTGTGGTGG 538
Db 411 AGCACGGGAGTGTCTTCTGAGACACCTTAAAGCCCATGCGGAAGAGAGAGCCCTTAC 470
Qy 539 GTTAAAGAAAAGACACAGTGGGCAACATTTGTATCGCGGTTCTACACCCGAAAGGAT 598
Db 471 ACCAAGGAAAAGACACAGTGGGACCTTCAAGCCCATGCGGAAGAGAGAGCCCTTAC 530
Qy 599 GTCCGGAGACATGGTGGTGCACACTGGAAGAAAGAGTCTCTCTGTCAGTATTGTGCA 658
Db 531 GTCCGACGCCACCTGGTGGTCCACACAGGATGCAAGGACTTCTGTGCCAGTTCTGTGCC 590
Qy 659 CAGAGATTTGGCGGAAAGGATCACTGACATATGAAGAAGAGTCAACATCAAGAG 718
Db 591 CAGAGATTTGGCGGCAAGGTTTCACTCACCCGGCATACCAAGAGAGCCCTCTCAGAG 650
Qy 719 CTCTGAGGTCAAAACAGAACAGTGGATTTCTCTGACCCATT 762
Db 651 CTGATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGCACCTT 694

RESULT 15

AAZ33549
ID AAZ33549 standard; cDNA; 2738 BP.
XX AC AAZ33549;
XX
XX 08-DEC-1999 (first entry)
XX
XX Human breast tumour-associated EST 9.
XX
XX Expressed sequence tag; EST; human; breast; cancer; cytostatic;
XX medicaments; gene therapy; treatment; fat metabolism; ss.
XX
XX Homo sapiens.
OS
XX DE19813835-A1.
XX
XX 23-SEP-1999.
PD
XX 20-MAR-1998; 98DE-01013835.
PF
XX 20-MAR-1998; 98DE-01013835.
PR
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA
XX Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
PI WPI; 1999-528979/45.
XX P-PSDB; AAY48468.
XX
XX Human nucleic acid sequences and protein products from normal breast
PT

PT tissue, useful for breast cancer therapy.

PS Claim 3; 96-97; 206pp; German.

This invention describes novel human nucleic acid sequences from normal breast tissue which have cytostatic activity. The nucleic acid sequences can be used to produce and isolate full-length gene sequences. They can be used to express proteins, which can be used as tools to find an activity against breast cancer. The sequences can be used in sense or antisense form. They are especially useful for medicaments for gene therapy to treat breast cancer and for treating illnesses associated with fat metabolism. AA233541-233610 represent expressed sequence tags described in the method of the invention

SQ Sequence 2738 BP; 771 A; 661 C; 546 G; 760 T; 0 U; 0 Other;

Query Match	Score 276;	DB 2;	Length 2738;
Best Local Similarity	18.4%		
Best Local Similarity	73.2%		

Best Local Similarity 71.2%; Pred. NO. 2.9e-76;
Matches 379; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 234 AATTCAAGGCACATGGGTACTCATTTCTCTGAGAAAAACCCACAAGTGTAAATTATGTG 293

Db

324 AATCAGTAGGCAAAATTGGCTACCCATTCTCCCGAAATCTCACCAGTGTGCTCACTGTG 383

QY
294 AGAAATGTTTCACCGGAAGATCATCTGAAGAATCACTCCATACACAGACCTAACA 353

Db
384 AGAGACGTTCAACCGGAAAGACCACTGAAAAACCACTTCAGACCCAGCCCAACA 443

QY 354 AAGAGACGTTTAAGTCGGAAGAAATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAAC 413

db
444 AAATGGCCTTTGGGTGTGAGGAGTGTGGGAGAGTACAACACCATGCTGGGTATAAG 503

414 GTCAC TTGGCC TTGCATG CCGCAACAAGTGGTGACCTCACCTGTAAGGTATGTTTGCAA 473

db 504 GGACCTGGCCCTCATGGGCAGCAGTGGGACCTACCTGTGGGGTCTGTGCCCTGG 563

474 CTTTGTAAAGCACGGGAGTGTCTTGAGCACCTTAAATCTCATGC---AGSCAAGTCGT 530

564 AGCTAGGGGAGCACCGAGGTGCTACTGGACCACTCAAGCCCCATCGGAGAGAGCCCC 623

531 CUGTGGGGGTTTAAAGAAAAAAGCACCAGTCGGAACATTGTGATCGCCGGTCTACACC 590

624 CTAGCGGAACCAAGGAAAAAGAAAGCAACAGTTCGGGACCTCTCTTAAAGTACGTCGTT

591 GAAGGATGTCGGAGACATGGTGGTGCACTGGGAAGAAAGCACTTCCCTCTCTCCACTC
592 CTTCCGGATCCCAAGGAAAAAGAACACCAAGTGCAGTCCGACCACTGTGAAAAGATGCTTCTACACC

684 GGAAGGATGTGGCAGCCCACTGGTGGTGTCCACACAGGATGCAAGGACCTTCTGTGCAGT 743

651 ATTGTGCACAGAGATTTGGCGGAAGGATTCACCTCACTCGACATATGAGAAGAGTCACA 710

db 744 TCTGTGCCACAGATTTGGCGCAAGGATCACCTACCCGGCATACCAAGAAGACCCACT 803

711 ATCAAGAGCTTCTGAAGGTCAAACAGAACCTGGATTTCTCTTGACCAATT 762

804 CACAGAGCTGATGAACAGAGAGCTTCAGACCGAGACCTTCTGAGACCTT 855

Search completed: November 8, 2004, 19:16:59

Job time : 790 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2004, 10:22:19 ; Search time 5072 Seconds
(without alignments)
10783.919 Million cell updates/sec

Title: US-09-242-772-116_COPY_480_1980

Perfect score: 1501

Sequence: 1 gatggccactgtcattctcg.....gtttccatcaagcttttccag 1501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1498.4	99.8	1503	9	AY420150	Homo sapi
2	1455.4	97.0	1503	9	AY420151	Pan trogl
3	1260	83.9	1500	9	AY420152	Mus muscu
4	805.8	53.7	1185	5	BX424854	BX424854
5	793.6	52.9	864	5	BX370812	BX370812
6	662.2	44.1	736	1	AL036879	AL036879
7	543.4	36.2	734	6	CA411972	CA411972
8	434	28.9	432	8	B94684	B94684
9	421.6	28.1	642	5	B210438	B210438
10	396.2	26.4	683	6	BY179727	BY179727
11	388.2	25.9	4767	3	AK040626	AK040626
12	373.8	24.9	1047	5	BX402039	BX402039
13	352.2	23.5	365	8	AQ125617	AQ125617
14	347	23.1	804	5	B260281	B260281
15	346.2	23.1	386	8	B99080	B99080
16	333.8	22.2	807	5	BUI186849	BUI186849
17	332.4	22.1	844	5	CN533596	CN533596
18	325.8	21.7	845	5	BUI72561	BUI72561
19	312.6	20.8	743	7	CN537373	CN537373
20	311	20.7	786	7	CK637889	CK637889
21	308.6	20.6	553	7	CN534810	CN534810
22	304.8	20.3	587	7	CN538026	CN538026
23	300.4	20.0	797	9	CN503111	CN503111
24	300.2	20.0	546	7	CN312865	CN312865

25	299.6	20.0	822	4	BI820689	BI820689 603034475
26	293.8	19.6	901	5	BUI161452	BUI161452 AGENCOURT
27	290.4	19.3	999	9	CNS03809	AL232866 Tetraodon
28	288.8	19.2	667	5	BX642593	BX642593 DKFZP686N
29	281	18.7	1098	1	AL523896	AL523896 AL523896
30	279.2	18.6	913	1	AL542380	AL542380 AL542380
31	278.2	18.5	854	1	AUI136661	AUI136661 AUI136661
32	273.2	18.2	720	7	CN312867	CN312867 170004245
33	271.8	18.1	617	7	CN665887	CN665887 A0832806-
34	266.8	17.8	563	6	CA750971	CA750971 UI-M-ROO-
35	266	17.7	838	7	CK601489	CK601489 AGENCOURT
36	262	17.5	711	7	CN392814	CN392814 170005315
37	261.8	17.4	600	4	BI989866	BI989866 4052-32 M
38	256.4	17.1	582	2	AW952782	AW952782 EST364852
39	250.2	16.7	700	5	BP761624	BP761624 BP761624
40	249.2	16.6	668	7	CF360369	CF360369 821696 MA
41	248.4	16.5	669	6	CB017609	CB017609 pgnic.pk0
42	246.2	16.4	1075	5	BX356755	BX356755 BX356755
43	240.8	16.0	855	4	BI826943	BI826943 603075831
44	238.4	15.9	582	6	CA949670	CA949670 IQ23F04.Y
45	238.2	15.9	964	1	AL526858	AL526858 AL526858

ALIGNMENTS

RESULT 1
AY420150

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

gene

ORIGIN

Query Match

Best Local Similarity

Matches 1499; Conservative

99.8%; Score 1498.4; DB 9; Length 1503;

99.9%; Pred. No. 0;

Mismatches 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Indels 0; Gaps 0;

Db 1 ATGGCCACTGTCATTCCTCGTGTGATTTTGTGCAAGTAAGAGATACCAGAAAGTCCCTTCA 60
 QY 62 GGGAAACGTTAAGCGTGTGTGAACCAACCAAGAAAGAACTTTCTTCCCAACTGTGTAC 121
 Db 61 GGGAAACGTTAAGCGTGTGTGAACCAACCAAGAAAGAACTTTCTTCCCAACTGTGTAC 120
 QY 122 AAGGCTTTAAGCTGTGTGAGAAATTAAGGTTCACTCTACTCTCTACACAGGAGAGAGG 181
 Db 121 AAGGCTTTAAGCTGTGTGAGAAATTAAGGTTCACTCTACTCTCTACTCTCTACACAGGAGAGG 180
 QY 182 CCTACAGTGCATACACAGACTGCACCAAGGCTTTGTTTCTTAAGTACAAATTAACA 241
 Db 181 CCTACAGTGCATACACAGACTGCACCAAGGCTTTGTTTCTTAAGTACAAATTAACA 240
 QY 242 AGGCATCTGGTCTACTCAITCTCTCGAGAAACCCCAAGTGTAATATTTGAGAAATG 301
 Db 241 AGGCATCTGGTCTACTCAITCTCTCGAGAAACCCCAAGTGTAATATTTGAGAAATG 300
 QY 302 TTTCCCGGAAGATCACTGAAGATCACTCTACACAGAGCCCTAACAAAGAGAGG 361
 Db 301 TTTCCCGGAAGATCACTGAAGATCACTCTACACAGAGCCCTAACAAAGAGAGG 360
 QY 362 TTTAAGTGCAGAGATGTGGCAAGAACTACAAATACCAAGCTTGGATTAAAGCTCACTTG 421
 Db 361 TTTAAGTGCAGAGATGTGGCAAGAACTACAAATACCAAGCTTGGATTAAAGCTCACTTG 420
 QY 422 GCCTTCATCGCGCAAGTGTGACCTCACTGTAAGTATGTTTGCAGAACTTTTGA 481
 Db 421 GCCTTCATCGCGCAAGTGTGACCTCACTGTAAGTATGTTTGCAGAACTTTTGA 480
 QY 482 AGCAGGGAGTCTCTGGAGCACTTAATCTCATGCAAGCAAGTCTGTGGGGGT 541
 Db 481 AGCAGGGAGTCTCTGGAGCACTTAATCTCATGCAAGCAAGTCTGTGGGGGT 540
 QY 542 AAGAAAAAGACCAAGTGCAGAACTGTGATGCGCGGTCTTACACCCGAAAGGATGTC 601
 Db 541 AAGAAAAAGACCAAGTGCAGAACTGTGATGCGCGGTCTTACACCCGAAAGGATGTC 600
 QY 602 CGGAGACATGTGTGTGACACATGGAAGAAAGACTTCTCTGTCAAGTATGTGCACAG 661
 Db 601 CGGAGACATGTGTGTGACACATGGAAGAAAGACTTCTCTGTCAAGTATGTGTGCACAG 660
 QY 662 AGATTGGCGAAGAGTACCTGACCTGACATGAGAGAGTCAATCAAGAGCTT 721
 Db 661 AGATTGGCGAAGAGTACCTGACCTGACATGAGAGAGTCAATCAAGAGCTT 720
 QY 722 CTGAAGTCAAAACAGACAGTGGATTTCTTACCCATTTACCTGCAATGTGTGTG 781
 Db 721 CTGAAGTCAAAACAGACAGTGGATTTCTTACCCATTTACCTGCAATGTGTGTG 780
 QY 782 CCTATAAAGACGAGCTCCTTCCGGTGTGATGCTTACCTTCCAGTGAAGTCAATCAAG 841
 Db 781 CCTATAAAGACGAGCTCCTTCCGGTGTGATGCTTACCTTCCAGTGAAGTCAATCAAG 840
 QY 842 CCATTCACAACTTCCTGAGTAAACCTCTACACACTCCATTTCACTGTCAGAGC 901
 Db 841 CCATTCACAACTTCCTGAGTAAACCTCTACACACTCCATTTCACTGTCAGAGC 900
 QY 902 TCGGATCTGCCCAACCAATGATCAACTTTTACCTTTGGGAATGACATGCCCAATAGAT 961
 Db 901 TCGGATCTGCCCAACCAATGATCAACTTTTACCTTTGGGAATGACATGCCCAATAGAT 960
 QY 962 AFGGACATGTTTCTTCTTCAAAAGAAAGAGCCATTAAGGGGAAATTCAGAGTACCTTG 1021
 Db 961 AFGGACATGTTTCTTCTTCAAAAGAAAGAGCCATTAAGGGGAAATTCAGAGTACCTTG 1020
 QY 1022 TATGCAATTTCTATTCTGAAAAGAGAGCCATTAAGGGGAAATTCAGAGTACCTTG 1081
 Db 1021 TATGCAATTTCTATTCTGAAAAGAGAGCCATTAAGGGGAAATTCAGAGTACCTTG 1080
 QY 1082 ATGGAGTTAAGAGTGGCGTCCCTCTTATCCCAAGATTTCTCAAGCTGTCATCATCT 1141
 Db 1081 ATGGAGTTAAGAGTGGCGTCCCTCTTATCCCAAGATTTCTCAAGCTGTCATCATCT 1140

QY 1142 AAGCTAGGTTGGATCTCTAGATGGTCCCTAGATGATGGTGCAGAGAGCTCTCCCTA 1201
 Db 1141 AAGCTAGGTTGGATCTCTAGATGGTCCCTAGATGATGGTGCAGAGAGCTCTCCCTA 1200
 QY 1202 TCCAAAAGCTCTATCTCCATCAGTGACCCCTAAACACACAGGATTTGATTTCTCAG 1261
 Db 1201 TCCAAAAGCTCTATCTCCATCAGTGACCCCTAAACACACAGGATTTGATTTCTCAG 1260
 QY 1262 TTGTTTAAATTTTATACCTTTAAATGGTCTCTCTATATCTCTATCAGTGGGAGCTT 1321
 Db 1261 TTGTTTAAATTTTATACCTTTAAATGGTCTCTCTATATCTCTATCAGTGGGAGCTT 1320
 QY 1322 GGAATGAGCTATTCACAGGAGAGACATTTCTTGTTCCTCAGCTCCCAACACAA 1381
 Db 1321 GGAATGAGCTATTCACAGGAGAGACATTTCTTGTTCCTCAGCTCCCAACACAA 1380
 QY 1382 CAGGATCTTCAGGATCTCTGCAAACTATAGGCTTGGTCTCTGCACTCAGTCTCAG 1441
 Db 1381 CAGGATCTTCAGGATCTCTGCAAACTATAGGCTTGGTCTCTGCACTCAGTCTCAG 1440
 QY 1442 GCTTTCCACAGAGTTTAAAGCAAGTACCACTCCCTCCAGCTTCCATCAAGCTTTTTCAG 1501
 Db 1441 GCTTTCCACAGAGTTTAAAGCAAGTACCACTCCCTCCAGCTTCCATCAAGCTTTTTCAG 1500

RESULT 2
 AY420151 1503 bp DNA linear GSS 17-DEC-2003
 LOCUS Pan troglodytes PLAG1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY420151 GI:39776108
 VERSION AY420151.1
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 (bases 1 to 1503)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 1503)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT This sequence as made by sequencing genomic exons and ordering them
 based on alignment.
 FEATURES
 source
 1..1503
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>1503
 /gene="PLAG1"
 /locus_tag="HCM7128"
 gene
 ORIGIN
 Query Match 97.0%; Score 1455.4; DB 9; Length 1503;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1456; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 QY 2 ATGGCCACTGTCATTCCTCGTGTGATTTTGTGCAAGTAAGAGATACCAGAAAGTCCCTTCA 61

Db 1 ATGGCCACTGTCAUCCCTGGTGAATTTGTCAAGATTAAGAGATACCCAGAAAGTCCCTTCA 60
 QY 62 GGGAAACGTAAAGCTGGTGAACCAACCAAGAAACCAATTTCTTGGCAACTGTGTGAC 121
 Db 61 GGGAAACGTAAAGCTGGTGAACCAACCAAGAAACCAATTTCTTGGCAACTGTGTGAC 120
 QY 122 AAGSCCTTTAAACAGTGTGTGAGAAATTAAGAGTTCACTCCCTACTCTCACAGGAGAGG 181
 Db 121 AAGSCCTTTAAACAGTGTGTGAGAAATTAAGAGTTCACTCCCTACTCTCACAGGAGAGG 180
 QY 182 CCCTAAGTGTATACAAACAGACTGACCAAGGCTTTTGTCTTAAGTCAAAATTAACA 241
 Db 181 CCCTAAGTGTATACAAACAGACTGACCAAGGCTTTTGTCTTAAGTCAAAATTAACA 240
 QY 242 AGSCACATGCTACTCTTCTCTGAGAAACCAACCAAGTGTAAATTTATGTGAGAAATG 301
 Db 241 AGSCACATGCTACTCTTCTCTGAGAAACCAACCAAGTGTAAATTTATGTGAGAAATG 300
 QY 302 TTTACCGGAAAGATCATCTGAAGATCACTCCATACACACACCCCTTAACAAAGAGAG 361
 Db 301 TTTACCGGAAAGATCATCTGAAGATCACTCCATACACACACCCCTTAACAAAGAGAG 360
 QY 362 TTTAAGTGGAGAAATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAACGTCACCTG 421
 Db 361 TTTAAGTGGAGAAATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAACGTCACCTG 420
 QY 422 GCCTTGCAATGCGCAACAGTGGTGAACCTCACTGTAAAGTATGTTTGCAAACTTTTGA 481
 Db 421 GCCTTGCAATGCGCAACAGTGGTGAACCTCACTGTAAAGTATGTTTGCAAACTTTTGA 480
 QY 482 AGCAGGGAGTCTTCTTGAGCAGCACTTAAATCTCATGAGCAAGTCTGTGTGGGGTT 541
 Db 481 AGCAGGGAGTCTTCTTGAGCAGCACTTAAATCTCATGAGCAAGTCTGTGTGGGGTT 540
 QY 542 AAAGAAAAAAGACCAAGTGGCAAGCAATTTGTGATCGCCGGTCTACACCGAAGATGTC 601
 Db 541 AAAGAAAAAAGACCAAGTGGCAAGCAATTTGTGATCGCCGGTCTACACCGAAGATGTC 600
 QY 602 CGGAGACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTGAGTATGTGCACAG 661
 Db 601 CGGAGACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTGAGTATGTGCACAG 660
 QY 662 AGATTTGGGCAAGAGTCACTGACTCGACATATGAAGAGAGTCAACATCAAGAGCTT 721
 Db 661 AGATTTGGGCAAGAGTCACTGACTCGACATATGAAGAGAGTCAACATCAAGAGCTT 720
 QY 722 CTGAAGTCAAAAAGACCAAGTGGATTTCTTGAACCAATTTACCTGCAATGTCTGTG 781
 Db 721 CTGAAGTCAAAAAGACCAAGTGGATTTCTTGAACCAATTTACCTGCAATGTCTGTG 780
 QY 782 CCTATAAAGACAGCTCTTCCGGTATGTCCTTACCTCCAGTGAACCTGTTATCAAG 841
 Db 781 CANNNTAAAGACAGCTCTTCCGGTATGTCCTTACCTCCANNAGTCTTATCAAG 840
 QY 842 CCATTCACAACTTTGCAAGTAAACCTCTACAACTCCATTTAGTCACTGTCAGAGC 901
 Db 841 CCATTCACAACTTTGCAAGTAAACCTCTACAACTCCATTTAGTCACTGTCAGAGC 900
 QY 902 TCGGGATTCGCCCAACCAATGATCAACATTTTACCTTTGGGAATGATGCCCAATAGAT 961
 Db 901 TCGGGATTCGCCCAACCAATGATCAACATTTTACCTTTGGGAATGATGCCCAATAGAT 960
 QY 962 ATGGACACTGTTTCATCCCTCTACACCTTTCTTTCAAAATATCCGTTCACTCTCA 1021
 Db 961 ATGGACACTGTTTCATCCCTCTACACCTTTCTTTCAAAATATCCGTTCACTCTCA 1020
 QY 1022 TATGCAATTTCTATTCCTGAAAGAAACAGCAATTAAGGGGAAATGAGAGTTACCTG 1081
 Db 1021 TATGCAATTTCTATTCCTGAAAGAAACAGCAATTAAGGGGAAATGAGAGTTACCTG 1080
 QY 1082 ATGGAGTTCAAGAGTGGCGTGCCTCTTTCATCCCAAGATTTCTCAAGCATCGTCATCT 1141
 Db 1081 ATGGAGTTCAAGAGTGGCGTGCCTCTTTCATCCCAAGATTTCTCAAGCATCGTCATCT 1140

QY 1142 AAGCTAGGTTGGATCCTCAGATGGTCCCTAGATGATGGTGCAGGAGACTCTCCCTA 1201
 Db 1141 AAGCTAGGTTGGATCCTCAGATGGTCCCTAGATGATGGTGCAGGAGACTCTCCCTA 1200
 QY 1202 TCCAAAAGCTCTACTCTCCATCAGTACCCCTAAACACACAGCATTTGGATTTTCTCAG 1261
 Db 1201 TCCAAAAGCTCTACTCTCCATCAGTACCCCTAAACACACAGCATTTGGATTTTCTCAG 1260
 QY 1262 TTGTTAAATTTATACCTTTAAATGGTCCCTCTATTAATCTCTATCAGTGGGAGCTT 1321
 Db 1261 TTGTTAAATTTATACCTTTAAATGGTCCCTCTATTAATCTCTATCAGTGGGAGCTT 1320
 QY 1322 GGAATGAGCTATTTCCAGGAGAACACATTTCTTGTTCCTCCAGCTCCCAACAACA 1381
 Db 1321 GGAATGAGCTATTTCCAGGAGAACACATTTCTTGTTCCTCCAGCTCCCAACAACA 1380
 QY 1382 CAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGCA 1441
 Db 1381 CAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGCA 1440
 QY 1442 GCTTTCACAGCAGTTTAAAGCACAAGTACCACTCCAGCTCCCAAGCTTTTTCAG 1501
 Db 1441 GCTTTCACAGCAGTTTAAAGCACAAGTACCACTCCAGCTTTCCATCAAGCTTTTTCAG 1500

RESULT 3
 AY420152
 LOCUS
 DEFINITION
 Mus musculus PLAG1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION
 AY420152
 VERSION
 AY420152.1
 GI:39776109
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 1500)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 Adams M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL
 Science 302 (5652), 1960-1963 (2003)
 PUBMED
 14671302
 REFERENCE
 2 (bases 1 to 1500)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 Adams M.D. and Cargill,M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT
 This sequence is made by sequencing genomic exons and ordering them
 based on alignment.
 FEATURES
 source
 1. .1500
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>1500
 /gene="PLAG1"
 /locus_tag="HMC7128"

ORIGIN
 Query Match 83.9%; Score 1260; DB 9; Length 1500;
 Best Local Similarity 90.5%; Pred. No. 0;
 Matches 1357; Conservative 0; Mismatches 140; Indels 3; Gaps 1;
 QY 2 ATGGCCTGTCATCTCTGGTGAATTTGTCAGAGTAAAGATACCCAGAAAGTCCCTTCA 61
 |||||

1 ATGGCCACTGTCATTCTCGTGATTTGTGACAGATAGAGATACCAGAAAGCCCTTCA 60
 62 GGGAAACGTAAGCGTGTGTAACCAACCAAGAAAAAATTTCTTCCCTGCAACTGTGAC 121
 61 GGGAAACGTAAGCGTGTGTAAGCAACCAAGAAAAAATTTCTTCCCTGCAACTGTGAC 120
 122 AAGCCCTTTAAAGCTGTTGAGAAATTAAGGTTCACTCCCTACTCTCACACAGGAGAGAG 181
 121 AAGCCCTTTAAAGCTGTTGAGAAATTAAGGTTCACTCCCTACTCTCACACAGGAGAGAG 180
 182 CCTACAGTGCATACACAGACTGACCAAGGCTTTGTTTCTTAAGTCAAAATTAACA 241
 181 CCTACAGTGCATACACAGACTGACCAAGGCTTTGTTTCTTAAGTCAAAATTAACA 240
 242 AGGCACATGGCTACTCAATCTCTCTGAGAAACCCCAAGTGAATTAATTTGAGAAATG 301
 241 AGGCACATGGCTACTCAATCTCTCTGAGAAACCCCAAGTGAATTAATTTGAGAAATG 300
 302 TTTCACCGGAAGATCATCTGAGAAATCACTCCATACACAGCCCTAACAAAGAGAG 361
 301 TTTCACCGGAAGATCATCTGAGAAATCACTCCATACACAGCCCTAACAAAGAGAG 360
 362 TTTCACCGGAAGATCATCTGAGAAATCACTCCATACACAGCCCTAACAAAGAGAG 421
 361 TTTCACCGGAAGATCATCTGAGAAATCACTCCATACACAGCCCTAACAAAGAGAG 420
 422 GCCTTCATCGCCCAACAAAGTGTGACCTCACTGTAAGTATGTTTGCACAACTTTTGA 481
 421 GCCTTCATCGCCCAACAAAGTGTGACCTCACTGTAAGTATGTTTGCACAACTTTTGA 480
 482 AGCAGGAGTGTCTCTGAGACCTTAATCTCATGACGCAAGTCTGCTGGGGGTT 541
 481 AGCAGGAGTGTCTCTGAGACCTTAATCTCATGACGCAAGTCTGCTGGGGGTT 540
 542 AAGAAAAAAGCAGCAGTCCGCAATGTGATGCGCGGTTCTACACCCGAAAGGATGTC 601
 541 AAGAAAAAAGCAGCAGTCCGCAATGTGATGCGCGGTTCTACACCCGAAAGGATGTC 600
 602 CGGAGACACATGTTGTCATCTGGAAGAAAGACTTCTCTGTCAGTATGTCACAG 661
 601 CGGAGACACATGTTGTCATCTGGAAGAAAGACTTCTCTGTCAGTATGTCACAG 660
 662 AGATTTGGCGAAGAGATCACTGACATGAGAGAGTCAATCAAGAGCTT 721
 661 AGATTTGGCGAAGAGATCACTGACATGAGAGAGTCAATCAAGAGCTT 720
 722 CTGAAGGTCAAAACAGAACAGTGGATTTCTTCAACCATTTACTGCAATGTGTCGTG 781
 721 CTGAAGGTCAAAACAGAACAGTGGATTTCTTCAACCATTTACTGCAATGTGTCGTG 780
 782 CCTATAAAGACGAGCTCCTTCCGGTGTATGTCCTTACCTTCAGTGAATCTTATCAAG 841
 781 CCTATAAAGACGAGCTCCTTCCGGTGTATGTCCTTACCTTCAGTGAATCTTATCAAG 840
 842 CCATTCACAAACACTTTCAGTGAATCACTTCAACACTCCATTTTCACTCCATGAGAGC 901
 841 CCATTCACAAACACTTTCAGTGAATCACTTCAACACTCCATTTTCACTCCATGAGAGC 900
 902 TCGGATCTGCCCAACAAATGATCAACTTTTACCTTTGGGAATGATGATGCCAATAGT 961
 901 TCGGATCTGCCCAACAAATGATCAACTTTTACCTTTGGGAATGATGATGCCAATAGT 960
 962 ATGGAACATGTTTCACTCTCACCACCTTTCTTTCAATATCCGTTTCACTTCACTCA 1021
 961 ATGGAACATGTTTCACTCTCACCACCTTTCTTTCAATATCCGTTTCACTTCACTCA 1020
 1022 TATGCAATTTCTATCTCTGAAAGAACAGCCATTAAGGGGGAATTTGAGATTAAGT 1081
 1021 TATGCAATTTCTATCTCTGAAAGAACAGCCATTAAGGGGGAATTTGAGATTAAGT 1080
 1082 ATGGAATTAAGAGTGGCGTCCCTCTTCTATCCCAAGATTTCTCAAGCATGTCATCT 1141
 1081 ATGGAATTAAGAGTGGCGTCCCTCTTCTATCCCAAGATTTCTCAAGCATGTCATCT 1140
 1137

1142 AAGTAGGTTGGATCTCAGATTGGTCTCCTAGATGATGGTGAGAGAGCTTCCCTA 1201
 1138 AAGTAGGTTAGAGCTCAGATTGGTCTCCTAGATGATGGTGAGAGCTTCCCTG 1197
 1202 TCCAAAAGCTCTATCTCCTATCAGTACGACCCCTTAAACACACAGATTTGGATTTCTCAG 1261
 1198 TCAAAGAGCTCTATCTCCTATCAGTACGACCCCTCAGCACACAGATTTGGATTTCTCAG 1257
 1262 TTGTTTAAATTTCAATCCTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1321
 1258 TTGTTTAAATTTCAATCCTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1317
 1322 GAATGAGCTATTTCCAGAGAGAGACATCTTCTGTTTCCAGCTCCCAACACAAACA 1381
 1318 GGGATGAGCTATTTCCAGAGAGAGAGACATCTTCTGTTTCCAGCTCCCAACACAAACA 1377
 1382 CAGGATCTTCCAGATCTCTGCAAACTATAGGCTTGGTCTCTGCACTCAGTCTCAGCA 1441
 1378 CAGGATCTTCCAGATCTCTGCAAACTATAGGCTTGGTCTCTGCACTCAGTCTCAGCA 1437
 1442 GCTTTCCAGCAGGTTTAAAGCAAGTACACCTCCCACTTCCATCAAGCTTTTTCAG 1501
 1438 GCGTTCCAGCAGGCTGAGCTCAAGCTTACCTTCCCTCCCTTCCAGGCGTTTTCAG 1497

RESULT 4
 BX424854 1165 bp mRNA linear EST 01-MAY-2004
 LOCUS
 DEFINITION XCL0BB001ZD02 5-PRIME, mRNA sequence.
 ACCESSION BX424854.2 GI:46928974
 VERSION BX424854.2
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 13, 2003 this sequence version replaced gi:30647825.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to this cluster, see
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?S=XCL0BB001ZD02RPL&c=4473.r.
 Location/Qualifiers
 1. .1165
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="XCL0BB001ZD02"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched, into
 double-strand cDNA was digested with Not I and cloned
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

FEATURES
 source
 1. .1165
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="XCL0BB001ZD02"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched, into
 double-strand cDNA was digested with Not I and cloned
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN
 Query Match 53.7%; Score 805.8; DB 5; Length 1165;
 Best Local Similarity 97.1%; Pred. No. 2.4e-225;
 Matches 871; Conservative 11; Mismatches 8; Indels 7; Gaps 6;

```
QY 235 ATTACAAAGGCACATGCTACTCATTTCTCTCTGAGAAAACCCCAAGTGTAATTATTGTGA 294
Db 275 AGTAATCAGCAGCATGCTACTCATTTCTCTCTGAGAAAACCCCAAGTGTAATTATTGTGA 334
QY 295 GAAATATGTTTACCGGAAAGATCATCTGGAAGATCAGCTCCATACACACGACCTTAACAA 354
Db 335 GAAATATGTTTACCGGAAAGATCATCTGGAAGATCAGCTCCATACACACGACCTTAACAA 394
QY 355 AGAGAGCTTTTAAGTGGGAAGATGTCGCAAGAACTACAATACCAAGCTTGGATTTAAACG 414
Db 395 AGAGAGCTTTTAAGTGGGAAGATGTCGCAAGAACTACAATACCAAGCTTGGATTTAAACG 454
QY 415 TCACCTTGGCTTTCATGCGCGCAACAGTGTGACCTCAGCTGTAAGGATGTTTGAACAC 474
Db 455 TCACCTTGGCTTTCATGCGCGCAACAGTGTGACCTCAGCTGTAAGGATGTTTGAACAC 513
QY 475 TTTTGAAGCAGCGGAGTCTCTGAGGACACCTTAAATCTCATGCGAGCAAGTCTGCTGG 534
Db 514 TTTTGAAGCAGCGGAGT-CTTCTGAGGACCTTAAATCTCATGCGAGCAAGTCTGCTGG 572
QY 535 TGGGCTTAAAGAAAAGACAGCAGTGGCAACATTTGTATCGCGGTTCTACACCCGAAA 594
Db 573 TGGGCTTAAAGAAAAGACAGCAGTGGCAACATTTGTATCGCGGTTCTACACCCGAAA 632
QY 595 GGATGTCGGGACACATGCTGTCACACTGGAGAAAGAGACTTCTCTGTGAGTATG 654
Db 633 GGATGTCGGGACACATGCTGTCACACTGGAGAAAGAGACTTCTCTGTGAGTATG 692
QY 655 TGCACAGAGATTTGGCGGAAAGGATCAGCTGACTCGACATATGAGAGAGAGTCAACATCA 714
Db 693 TGCACAGAGATTTGGCGGAAAGGATCAGCTGACTCGACATATGAGAGAGAGTCAACATCA 752
QY 715 AGAGCTTCTGAAGTCAAAAAGACAGACAGTGGATTTCTTGAACCATTTTACCTGCAAGT 774
Db 753 AGAGCTTCTGAAGTCAAAAAGACAGACAGTGGATTTCTTGAACCATTTTACCTGCAAGT 812
QY 775 GTCTGTGCTATAAAGACGAGCTCTTCCGGTATGTCCTTACCTCCAGTGAAGTGT 834
Db 813 GTCTGTGCTATAAAGACGAGCTCTTCCGGTATGTCCTTACCTCCAGTGAAGTGT 872
QY 835 ATCAAGGCCATTACAAACACTTTTGCAGTTAAACCTCTACAACTCCATTTTCAAGTCCAT 894
Db 873 ATCAAGGCCATTACAAACACTTTTGCAGTTAAACCTCTACAACTCCATTTTCAAGTCCAT 932
QY 895 GCAGAGCTCGGATCTGCCACCAAAATGATCAAACTTTTACCTTTGGGAATGACATGCC 954
Db 933 GCAGAGCTCGGATCTGCCACCAAAATGATCAAACTTTTACCTTTGGGAATGACATGCC 992
QY 955 AATAGATATGACATGTTTCAATCCC-TCTCACCACCTTTCTTTCAAAATATCCGTTCAAGT 1013
Db 993 AATAGATATGACATGTTTCAATCCC-TCTCACCACCTTTCTTTCAAAATATCCGTTCAAGT 1052
QY 1014 CTACCTCATATGCAATTTCTATTCCTGAAAAGAAAGACAGCCATTAAGGGGGAATTCAGA 1073
Db 1053 CTACCTCATATGCAATTTCTATTCCTGAAAAGAAAGACAGCCATTAAGGGGGAATTCAGA 1110
QY 1074 GTTACCTGATGAGTACAGGTGGGCGCTCTTTCATCCCAAGATTTCTCAAGCAT 1130
Db 1111 RKTACCGGATGGAGTACAGKGGC-TKCCCTCTTCATCCC-ARATWCTCAAGCAT 1165
```

```
RESULT 5
BX370812 864 bp mRNA linear EST 26-APR-2004
LOCUS BX370812 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL012YM24 5-PRIME, mRNA sequence.
ACCESSION BX370812
VERSION BX370812.2
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE

1 (bases 1 to 864)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 8, 2003 this sequence version replaced gi:30445558.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4473.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAG062D02_CS05868_1&c=4473.r

FEATURES

Location/Qualifiers
1..864
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL012YM24"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_lines="RAMOS CELL LINE"
/clone_lib="HOMO sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match	Best Local Similarity	Score	793.6;	DB 5;	Length 864;
Matches	843;	Conservative	0;	Mismatches	9;
				Indels	4;
				Gaps	4;
QY	505	CCCTTAATCTCATGAGGCAAGTCTCTGTTGGGTTAAAGAAAAGACACCAAGTGGCA	564		
Db	13	CCCTTAATCTCATGCA-GCAAGTCTCTGTTGGGTTAAAGAAAAGC-CCAGTGGCA	70		
QY	565	ACATTGTGATGCGCGTTCTACCCGAAAGGATGTCGGAGACACATGTTGGTGCACAC	624		
Db	71	ACATTGTGATGCGCGTTCTACCCG-AAGGATGTCGGAGACACATGTTGGTGCACAC	129		
QY	625	TGGAAGAAAGGACTTCTCTGTGATGTCGACAGAGATTTGGGCGAAGGATACCT	684		
Db	130	TGGAAGAAAGGACTTCTCTGTGATGTCGACAGAGATTTGGGCGAAGGATACCT	189		
QY	685	GATCGACATATGAAGAAGAGTCACAATCAAGAGCTTCTGAAGTCAAAACAGACCACT	744		
Db	190	GATCGACATATGAAGAAGAGTCACAATCAGGAGCTTCTGAAGTCAAAACAGACCACT	249		
QY	745	GGATTTCCTTGACCAATTTACCTGCAATGTTGTGCTTATATAAAGAGAGAGCTCTTCC	804		
Db	250	GGATTTCCTTGACCAATTTACCTGCAATGTTGTGCTTATATAAAGAGAGAGCTCTTCC	309		
QY	805	GGTGATGTCCTTACCTTCCAGTCACTGTTATCAAGCCATTCAAAAACACTTTGAGTT	864		
Db	310	GGTGATGTCCTTACCTTCCAGTCACTGTTATCAAGCCATTCAAAAACACTTTGAGTT	369		
QY	865	AAACCTCTACACACTCCATTTTCACTTCAGTCCATGACAGAGCTCGGATCTGCCACCAATGAT	924		
Db	370	AAACCTCTACACACTCCATTTTCACTTCAGTCCATGACAGAGCTCGGATCTGCCACCAATGAT	429		
QY	925	CACAACTTTACCTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCATCCCTCTCA	984		
Db	430	CACAACTTTACCTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCATCCCTCTCA	489		
QY	985	CCACCTTTCTTCAATATCCGTTTCACTTCTACCTCATATGCAATTTCTATTCCTGAAAA	1044		

490 CCACCTTTCTTCAAAATATCCGTTTCAGTTCTACCTCATATGCAATTTCTATTCTCCTGAAA 549
1045 AGAACAGCCATTAAAGGGGGAATTTGAGAGTTACCTGATGAGTTACAGGTGCGTGCC 1104
550 AGAACAGCCATTAAAGGGGGAATTTGAGAGTTACCTGATGAGTTACAGGTGCGTGCC 609
1105 CTCTTATCCCAAGATTTCAAGCATGTCATCATCTAAGCTAGGTTGGATCCTCAGAT 1164
610 CTCTTATCCCAAGATTTCAAGCATGTCATCATCTAAGCTAGGTTGGATCCTCAGAT 669
1165 TGGTCCCTAGATGATGTCGAGGAGCTCTCCCTATCCAAAAGCTCTATCTCCATCAG 1224
670 TGGTCCCTAGATGATGTCGAGGAGCTCTCCCTATCCAAAAGCTCTATCTCCATCAG 729
1225 TGACCCCTAAACACACACAGCATTTGGATTTTCTCAGTGTGTTAATTTACCTTTAAA 1284
730 TGACCCCTAAACACACAGCATTTGGATTTTCTCAGTGTGTTAATTTACCTTTAAA 789
1285 TGGTCCCTAAATATCTCTATCAGTGGGAGCTTGAATGAGTATTTCCAGGAAGA 1344
790 TGGTCCCTAAATATCTCTATCAGTGGGAGCTTGAATGAGTATTTCCAGGAAGA 848
1345 AGCAGATCTCTGTT 1360
849 AGCAGATCTCTGTT 864

RESULT 6
AL036879 736 bp mRNA linear EST 04-SEP-2003
LOCUS DKFp564P1863_r1 564 (synonym: hfb2) Homo sapiens cDNA clone
DEFINITION DKFp564P1863 5', mRNA sequence.
ACCESSION AL036879
VERSION AL036879.3 GI:5927934
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
EST (Duesterhoeft, et al.)
Unpublished (1999)
On Jul 7, 1999 this sequence version replaced gi:5866332.
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de.
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
No si sequence available.
This clone (DKFp564P1863) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFp564P1863"
/tissue.type="brain"
/dev.stage="fetal"
/lab_host="X1-2blue"
/clone.lib="564 (synonym: hfb2)"
/notes="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

FEATURES
source
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFp564P1863"
/tissue.type="brain"
/dev.stage="fetal"
/lab_host="X1-2blue"
/clone.lib="564 (synonym: hfb2)"
/notes="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

ORIGIN
Query Match 44.1%; Score 662.2; DB 1; Length 736;
Best Local Similarity 98.1%; Pred. No. 3.9e-183;
Matches 723; Conservative 0; Mismatches 8; Indels 6; Gaps 5;

280 GTGTAAATTTGTGAGAAAATGTTTACCGGAAAGATCATCTGAAGAATCACCTCCATPAC 339
2 GGGTAATTTATTTGTGAGAAAATGTTTACCGGAAAGATCATCTGAAGAATCACCTCCATPAC 61
340 ACAGACCCCTAAACAAAGAGACGTTTAAAGTGGGAAAGATTTGGGCAAGAACTCAATACCAA 399
62 ACAGACCCCTAAACAAAGAGACGTTTAAAGTGGGAAAGATTTGGGCAAGAACTCAATACCAA 121
400 GCTTGGATTTAAACGTCATCTTGGGCTTCATGTCGCAACAAAGTGGTGGTCACTCACTGTA 459
122 GCTTGGATTTAAACGTCATCTTGGGCTTCATGTCGCAACAAAGTGGTGGTCACTCACTGTA 181
460 GGTATGTTTGAACACTTTTGAAGCAGCGGAGTCTTCTGGAGCACCTTTAAATCTCATGC 519
182 GGTATGTTTGAACACTTTTGAAGCAGCGGAGTCTTCTGGAGCACCTTTAAATCTCATGC 241
520 AGGCAAGTCGTCGTGGGTTTAAAGAAAAAAGCACCAGTGCAGCAATTTGTGATGCGCG 579
242 AGGCAAGTCGTCGTGGTGTGTAAAGAAAAAAGCACCAGTGCAGCAATTTGTGATGCGCG 301
580 GTTCTACACCCGAAAGAGTGTCCGGAGACACATGTTGTGTCACACTTGGAGAAAGGACTT 639
302 GTTCTACACCCGAAAGAGTGTCCGGAGACACATGTTGTGTCACACTTGGAGAAAGGACTT 361
640 CCTCTGTGAGTATTTGTCACAGAGATTTGGGCGAAAGAGTACCTGACTCGACATATGAA 699
362 CCTCTGTGAGTATTTGTCACAGAGATTTGGGCGAAAGAGTACCTGACTCGACATATGAA 421
700 GAAGAGTCAATCAAGAGCTTCTGAAGTCAAAACAGAAACGAGTGGTTCCTTTCACCC 759
422 GAAGAGTCAATCAAGAGCTTCTGAAGTCAAAACAGAAACGAGTGGTTCCTTTCACCC 481
760 ATTTACCTGCAATGTTCTGTGTCCTATAAAGACGAGTCTCTCCGGTGTGATGCTTACCC 819
482 ATTTACCTGCAATGTTCTGTGTCCTATAAAGACGAGTCTCTCCGGTGTGATGCTTACCC 541
820 TTCCAGTG-AACTGTTTATCAAGCCATTCAAAACACTTTGAGTGTAAAGCTTCAACA 878
542 TTCCAGTGAAACTGTTTATCAAGCCATTCAAAACACTTTGAGTGTAAAGCTTCAACA 601
879 CTCCTTTTCACTGTCATGC-AGAGTCCGGATTCGCCACC-AAATGATCAACTTTACC 936
602 CTCCTTTTCACTGTCATGC-AGAGTCCGGATTCGCCACC-AAATGATCAACTTTACC 661
937 TTTGGGAATGACATGCCCAATAGATAT-GGACATGTTTCCTCTCACCACCTTTCTT 995
562 TTTGGGAATGACATGCCCAATAGATATGGGACACTGTTCAATCTCTCAACAATTTCTT 721
996 TCAAAATCCGTTTCAGT 1012
722 --CAATATCCGTTTCAGT 736

RESULT 7
CA411972/c 734 bp mRNA linear EST 07-NOV-2002
LOCUS UI-H-EZ0-bal-m-01-0-UI.s1 NCI-CGAP_chi Homo sapiens cDNA clone
DEFINITION UI-H-EZ0-bal-m-01-0-UI 3', mRNA sequence.
ACCESSION CA411972
VERSION CA411972.1 GI:24774623
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 734)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of Orthopaedics
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
 The following repetitive elements were found in this cDNA sequence: 1-22, >AT-rich#low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Location/Qualifiers

1..734
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-EZO-bal-m-01-0-UI"
 /tissue_type="Chondrosarcoma Grade II"
 /dev_stage="Adult"
 /lab_host="PH10B (Life Technologies)"
 /clone_lib="NCI CGAP Ch1"
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP Ch1 is a cDNA library containing the following tissue(s): Chondrosarcoma Grade II. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is TGATCAGCT.
 TAG TISSUE=grade-2-chondrosarcoma
 TAG LIB=UI-H-EZO
 TAG_SEQ=ATCTAATATG"

ORIGIN

Query Match 36.2%; Score 543.4; DB 6; Length 734;
 Best Local Similarity 99.5%; Pred. No. 3.7e-148;
 Matches 555; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 QY 945 TGACATGCCCAATAGATGACACTGTTTCATCCCTCTCACCACCTTTCTTCAAATATC 1004
 Db 734 TGACATGCCCAATAGATGACACTGTTTCATCCCTCTCACCACCTTTCTTCAAATATC 675
 QY 1005 CGTTTCAGTTCTACTCATATGCAATTTCTATTCTCTGAAAAGAACAGCCATTAAAGGG-G 1063
 Db 674 CGTTTCAGTTCTACTCATATGCAATTTCTATTCTCTGAAAAGAACAGCCATTAAAGGGNN 615
 QY 1064 GAAATTGAGAGTTTACCTGATGGAGTTACAAGTGGGTGCGCTTTCATCCCAAGATTCT 1123
 Db 614 GAAATTGAGAGTTTACCTGATGGAGTTACAAGTGGGTGCGCTTTCATCCCAAGATTCT 555
 QY 1124 CAAGCATCTCATCATCTAAGCTAGGTTGAGTTCCTCAGATTGGGTCCCTTAGATGATGGT 1183
 Db 554 CAAGCATCTCATCATCTAAGCTAGGTTGAGTTCCTCAGATTGGGTCCCTTAGATGATGGT 495
 QY 1184 CGAGGAGACTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACCA 1243
 Db 494 CGAGGAGACTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACCA 435
 QY 1244 GAATTGGATTTTCTCAGTTGTTTAAATTTTCAATACCTTTAAATGGTCCCTTATATCT 1303
 Db 434 GAATTGGATTTTCTCAGTTGTTTAAATTTTCAATACCTTTAAATGGTCCCTTATATCT 375
 QY 1304 CTATCAGTGGGAGCCTTGGATAGCTATTCCAGGAAAGAGCACATCTTCTGTTTCC 1363
 Db 374 CTATCAGTGGGAGCCTTGGATAGCTATTCCAGGAAAGAGCACATCTTCTGTTTCC 315
 QY 1364 CAGTCCCCACACAAACAGAGATCTTCAGGATCTTCAGGATCTCGAAACACTATAGGGCTTGGGTCT 1423

Db 314 CAGTCCCCCAACAAACACAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTCAGGATCT 255
 QY 1424 CTGCATCTCACTGTCAAGAGCTTTTCCAGCAGTTTAAAGCAAGTACACCTCCACAGT 1483
 Db 254 CTGCATCTCACTGTCAAGAGCTTTTCCAGCAGTTTAAAGCAAGTACACCTCCACAGT 195
 QY 1484 TTCCATCAAGCTTTTTCAG 1501
 Db 194 TTCCATCAAGCTTTTTCAG 177

RESULT 8

B94684/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: CIT-HSP-2170018.TR

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: M13-21;

Class: BAC ends.

Location/Qualifiers

1..492

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="GDB:7103260"

/db_xref="taxon:9606"

/clone="2170018"

/sex="Male"

/cell_type="Sperm"

/clone_lib="CIT-HSP"

/note="Vector: pBelobAC11; Site 1: HindIII; Site 2: HindIII"

ORIGIN

Query Match 28.9%; Score 434; DB 8; Length 492;

Best Local Similarity 98.9%; Pred. No. 5.3e-116;

Matches 437; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1055 TTAAAGGGGAAATTGAGAGTTTACCTGATGGAGTTACAGAGTGGGCGTCCCTTTCATCC 1114

Db 492 TTAAAGGGGAAATTGAGAGTTTACCTGATGGAGTTACAGAGTGGGCGTCCCTTTCATCA 433

QY 1115 CAAGATTCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCAGATTGGTCCCTTA 1174

Db 432 CAAGATTCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCAGATTGGTCCCTTA 373

QY 1175 GATGATGTGAGGAGACCTCTCCCTATCCAAAAGCTCTATCTCCAGTACCCCTTA 1234

Db 372 GATGATGTGAGGAGACCTCTCCCTATCCAAAAGCTCTATCTCCAGTACCCCTTA 313

(1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN	
Query Match	28.1%; Score 421.6; DB 5; Length 642;
Best Local Similarity	86.0%; Pred. No. 2.6e-112;
Matches	502; Conservative 0; Mismatches 79; Indels 3; Gaps 3;
QY	209 ACCAAGGCGCTTTGTTCTTAAGTACAAATTACAAAGGCATGGTACTCATCTCTCTGAG 268
DB	41 ACCAAATCTCAGGACCTTTAGTGGGCAACATCAGGCATATGGCTACTCATCTCTCTGAG 100
QY	269 AAAACCCACAGTGTAAATTTTGTGAGAAATGTTTACCGGAAAGATCATCTGAAGAT 328
DB	101 AAAACCCACAGTGTAAATTTTGTGAGAAATGTTTACCGGAAAGATCATCTGAAGAT 159
QY	329 CACCTCCATACACACGACCTTAACAAGAGAGAGCTTTAAGTGGGAAGATGGGCAAGAC 388
DB	160 CACCTACATACACAATCCCAAGAGAGGCTTTAAGTGTGAAGATGTGGAAGAAC 219
QY	389 TACAATACCAAGCTTGGATTTAAACGTCTTGGCTTGCATGCCGCAACAAGTGTGTGAC 448
DB	220 TACAATACCAAGCTTGGATTTAAACGTCTTGGCTTGCATGCCGCAACAAGTGTGTGAC 279
QY	449 CTCACCTGTAAAGTATGTTTGCAAACTTTGAAAGCACGGAGTGTCTCTGGAGCACCTT 508
DB	280 CTCACCTGTAAAGTATGTTTGCAAACTTTGAAAGCACGGAGTGTCTCTGGAGCACCTT 339
QY	509 AAATCTCATGACGAGCAAGTCGTCTGGTGGGTTAAAGAAAAGACACAGTGGCAACAT 568
DB	340 AAATCTCATGACGAGCAAGTCGTCTGGTGGGTTAAAGAAAAGACACAGTGGCAACAT 399
QY	569 TGTGATTCGCGGTTTACACCGAAGAGTGTCCGAGACACATGTGTGTGACACCTGGA 628
DB	400 TGTGATTCGCGGTTTACACCGAAGAGTGTCCGAGACACATGTGTGTGACACCTGGA 459
QY	629 AGAAGGACCTTCTCTGTCTAGTATTTGACACAGATTTGGGCGAAGGATCACCTGACT 688
DB	460 AGAAGGACCTTCTCTGTCTAGTATTTGACACAGATTTGGGCGAAGGATCACCTGACT 518
QY	689 CGACATATGAAGAGAGTCAACAATCAAGAGCTTCTGAAGTC-AAAACAGACCAAGTGA 747
DB	519 CGCCACATGAAGAGAGTCAACAATCAAGAGCTTCTGAAGTC-AAAACAGACCAAGTGA 578
QY	748 TTTCCTTGACCCATTTACCTGCATGTGTCTGTGCTTATTAAG 791
DB	579 CCTCTAGATCCCTTTTACCTGCAAGCTTCTGTGCTTATCCAGG 622

RESULT 10
BY719727
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BY719727 683 bp mRNA linear EST 17-DEC-2002
BY719727 RIKEN full-length enriched, in vitro fertilized eggs Mus
musculus cDNA clone 7420412G02 5', mRNA sequence.

BY719727.1 GI:27132844
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 683)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Sult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalow, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustigindich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,

QY	1235 AACACACAGCATGGATTTTCTCAGTGTGTTAAATTTATACCTTTAAATGGTCTCTCCC 1294
DB	312 AACACACAGCATGGATTTTCTCAGTGTGTTAAATTTATACCTTTAAATGGTCTCTCCC 253
QY	1295 TATAATCTCTATCAGTGGGAGCCTTGGATGAGCTATCCCAAGGAAGAGACATCT 1354
DB	252 TATAATCTCTATCAGTGGGAGCCTTGGATGAGCTATCCCAAGGAAGAGACATCT 193
QY	1355 TCTGTTTCCAGCTCCCAACACACAGAGATCTTCCAGGATCTCTGCAAAACATATAGGG 1414
DB	192 TCTGTTTCCAGCTCCCAACACACAGAGATCTTCCAGGATCTCTGCAAAACATATAGGG 133
QY	1415 CTTGGGTCTCTGCACTGTCAGCAGCTTTCACAGCAGTTTAAGCAGATACCAACC 1474
DB	132 CTTGGGTCTCTGCACTGTCAGCAGCTTTCACAGCAGTTTAAGCAGATACCAACC 73
QY	1475 CTCCACAGTTTCCATCAAGCTT 1496
DB	72 CTCCACAGTTTCCATCAAGCTT 51

RESULT 9
BU210438
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

BU210438 642 bp mRNA linear EST 25-NOV-2002
604156514F1 CSEQCHN03 Gallus gallus cDNA clone ChEST100514 5', mRNA
sequence.
BU210438 GI:25383924
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 642)
Boardman, P.E., Sanz-Esquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .642
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST100514"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN03"
/note="Organ: whole embryo; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research

FEATURES
source

Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, I., Hori, F., Imoto, S., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Konno, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

TITLE

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

ORIGIN

Query Match 26.4%; Score 396.2; DB 6; Length 683;
 Best Local Similarity 87.1%; Pred. No. 8e-105;
 Matches 447; Conservative 0; Mismatches 63; Indels 3; Gaps 1;

QY 975 ATCCCTCTCACCACTTTCTTTTCAATATCCCTTACCTCATATGCAATTCTTA 1034
 Db 1 ACCCTCTCATCATCTTGTCTTCAATGCCATTTCAGTTCTACCTCATACGCAATCTCTA 60

QY 1035 TTCTGTAAGAAAGACAGCCATTAAAGGGGGAATTTGAGAGTTACCTGATGGAGTTACAG 1094
 Db 61 TTCTGTAAGAAAGACAGCCATTAAAGGGGGAATTTGAGAGTTACCTGATGGAGTTACAG 120

QY 1095 GTGGGTGCTCTCTTCTTCAATCCCAAGATTCTCAAGCATCTCATCTAAGCTAGGTTGG 1154
 Db 121 GTGGTGCACCATCTTTCATCCAGGATCTCCAGCATCG---TCATCTAAGCTAGGTTGG 177

QY 1155 ATCTCTAGATTTGGTCCCTAGATATGTGTGAGAGAGCTCTCCCTATFCCAAAGCTCTA 1214
 Db 178 AGCTCTAGATTTGGTCCCTAGATATGTGTGAGAGAGCTCTCCCTATFCCAAAGCTCTA 237

QY 1215 TCTCATCATGTCAGCCCTTAAACACACAGCATTTGGATTTTCTCAGTTGTTTAATTCA 1274
 Db 238 TCTCATCATGTCAGCCCTTAAACACACAGCATTTGGATTTTCTCAGTTGTTTAATTCA 297

QY 1275 TACCTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCTTTGGATGAGCTATT 1334
 Db 298 TACCATTTGAATGGTCCCTTATAACCCACTTTTCAGTGGGAGCTTTGGATGAGCTATT 357

QY 1335 CCCAGGAAGAGACATTTCTTCTGTTCAGCTCCCCACAAACACAGGATCTTCAGG 1394
 Db 358 CCCAGGAAGAGACATTTCTTCTGTTCAGCTCCCCACAAACACAGGATCTTCAGG 417

QY 1395 ATCTGCAACACTATAGGGCTTTGGTCTCTGACACTCTGTCTCAGAGCTTTCCAGCA 1454
 Db 418 ACCCTGCAACACTGTGGGTCTTAAGTTCTCTGACACTCTGTCTCAGAGCTTTCCAGCA 477

QY 1455 GTTTAAGCAAGTACCAACCTCCACGTTTCC 1487
 Db 478 GCTTGAGCTCAAGCACTACCTTGCCCCGTTTCC 510

RESULT 11

AK040626

LOCUS

DEFINITION

MUS MUSCULUS 0 DAY NEONATE THYMUS CDNA, RIKEN FULL-LENGTH ENRICHED

LIBRARY, CLONE:A430110122 PRODUCT:PLEIOMORPHIC ADENOMA GENE-LIKE 2,

FULL INSERT SEQUENCE.

ACCESSION AK040626

VERSION AK040626.1

KEYWORDS HTG; CAP TRAPPER.

SOURCE MUS MUSCULUS (HOUSE MOUSE)

ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

FEATURES

source

Location/Qualifiers

1..683

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="7420412602"

/sex="female"

/tissue_type="in vitro fertilized eggs"

/dev_stage="egg"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, in vitro

fertilized eggs"

/note="Site 1: Sali; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

QY 818 CCTTCCAGTGAACGTGTT---ATCAAGCCATTCACAAACACTTTGCAAGTTAAACCTCTAC 874
 Db |||||
 QY 1039 GCCTCTCGGACGTGATGGGGGCTTAAGGCCCTTCTCGCATGTTGCCATGGGTATGTAT 1098
 Db |||||
 QY 875 AACACTCCATTTCAGTCCATGCAGAGCTCGGGATCTGCCCA---CCAAATGATCAAACT 931
 Db |||||
 QY 1099 GGTGCCACATCCCTACCATGCCAGTGGGGGCGATGCCACATCCCTCGGTGCAACAACA 1158
 Db |||||
 QY 932 TTACCTTTGGGAATGATGCCCAATAGATATGACACTGTTTCATCCCTCTCCACCACTT 991
 Db |||||
 QY 1159 CTGCCCATGGGTATGAGTTACCTCTGGAATCTTCTCTATCTCTCTCCTCAGCTCAGCTC 1218
 Db |||||
 QY 992 TCTTTCAAAATATCGTTTCAGTCTTACTCTATGATGATGCAATTTCTATCTCTGAAAAGAACAG 1051
 Db |||||
 QY 1219 CTTCCAAATACAGCTTGGATCTACCTCTA-----CTTGCCCGCAAA 1263
 Db |||||
 QY 1052 CCATTAAAGGGGAAATTGAGATTTACCTGATGAGTTACAGGTTCACAGGTGGCGCTCTTCA 1111
 Db |||||
 QY 1264 CTGCCCAAGTGGAGTGGATAGTTTCTGGCAGAGCTTCTGGAAAGCCTGTCTCTCTCC 1323
 Db |||||
 QY 1112 TCCCAAGATTTCAAGATCGTCAATCTAAGCTAGGTTGGATCTCTCAGATTTGGGTCC 1171
 Db |||||
 QY 1324 TCGGCTGAACCTCAGCCCGCTCA-----CCTCAGCCGGCGCA 1362
 Db |||||
 QY 1172 CTAGATGATGTTGCGAGGAGCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCCC 1231
 Db |||||
 QY 1363 GCTGAGCCCTCTGTGATGAAGCATGTCTGCGCAAGAGTCCGCCAACCTCTCCGAGGCC 1422
 Db |||||
 QY 1232 CTAAACACACCAAGCATTTGATTTTCTCAGTTTGTATTAATTTTCATACCTTTAATGGTCT 1291
 Db |||||
 QY 1423 CTCTGGCTGCTPAATGTGAGCTTTTCCCACTTACTGGGCTTTCTTCGCGTCAACCTACCC 1482
 Db |||||
 QY 1292 CCCTATATCTCTTA---TCAGTGGGAGCCCTTGAATGAGCTATTCACAGGAAAGCA 1348
 Db |||||
 QY 1483 CCATGTAACCCACAGGGGCCACAGAGGCTTGGTTATGGGCTACTCCCAAGCTGAGCA 1542
 Db |||||
 QY 1349 CATCTCTGTCTTCCAGCTCCCAACACAAACAGGATCTTCAGATCTTCGAAACT 1408
 Db |||||
 QY 1543 CAGCTTTGCTCACACTTTGAGGCTCAGGCTCAGATTCGCTGGAGCTGGGGGACC 1602
 Db |||||
 QY 1409 ATAGGGCTTGGGTCTGTCACCTGTCAGAGCTTTCACAGAGCTTTCACAGCAGTTAAGCACAAGT 1468
 Db |||||
 QY 1603 CTGAACCTTTGGGCTCTGCACTCTCTTACCTCTGTTTCACCTCTGGCCTGAGTACC--- 1659
 Db |||||
 QY 1469 ACCACCTTCCACAGTTTCCATCAAGCTTTTCCAG 1501
 Db |||||
 QY 1660 ACCACCTTGGCTCTGTTTCCACCAAGCATTCAG 1692
 Db |||||

RESULT 12

BX402039

LOCUS

DEFINITION

Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED

Homo sapiens cDNA clone CS0DL012YM24 5-PRIME, mRNA sequence.

BX402039

VERSION

BX402039.2

GI:46835543

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

Homo sapiens

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr

Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4473.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0DL012HBG12QP1&c=4473.r.

FEATURES

source

1. 1047

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DL012YM24"

/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"

/cell_line="RAMOS CELL LINE"

/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 24.9%; Score 373.8; DB 5; Length 1047;
 Best Local Similarity 96.0%; Pred. No. 3.6e-98;
 Matches 408; Conservative 6; Mismatches 8; Indels 3; Gaps 3;
 QY 239 CAAAGGCATCGGTACTCTCTCTGAGAAAACCCACAGTGTAAATTTGTGAGAA 298
 Db |||||
 QY 544 CAGAAGCATCGGTACTCTCTCTGAGAAAACCCACAGTGTAAATTTGTGAGAA 603
 Db |||||
 QY 299 ATGTTTCCCGGAAAGATCATCTGAAGATCACTCCATACACGACCCCTAACAAAGAG 358
 Db |||||
 QY 604 ATGTTTCCCGGAAAGATCATCTGAAGATCACTCCATACACGACCCCTAACAAAGAG 663
 Db |||||
 QY 359 ACGTTTAAAGTCGGAAGATGTGGCAAGACTACATACCAAGCTTGGATTTAAACGTAC 418
 Db |||||
 QY 664 ACGTTTAAAGTCGGAAGATGTGGCAAGACTACATACCAAGCTTGGATTTAAACGTAC 723
 Db |||||
 QY 419 TTGGCTTTGATGCGCAACAAAGTGGTGAACCTCACTTAAAGTATGTTTGCACACTTTT 478
 Db |||||
 QY 724 TTGGCTTTGATGCGCAACAAAGTGGTGAACCTCACTTAAAGTATGTTTGCACACTTTT 783
 Db |||||
 QY 479 GAAAGCAGGGAGTCTCTGG-AGCACCTTAAATCTCAGCAGCAGTCTGTGGTGG 537
 Db |||||
 QY 784 GAAAGCAGGGAGTCTCTGG-AGCACCTTAAATCTCAGCAGCAGTCTGTGGTGG 843
 Db |||||
 QY 538 GGTTAAGAAAAAAGCAACCAAGTGGTGAACCTCACTTAAAGTATGTTTGCACACTTTT 597
 Db |||||
 QY 844 GGTTAAGAAAAAAGCAACCAAGTGGTGAACCTCACTTAAAGTATGTTTGCACACTTTT 902
 Db |||||
 QY 598 TGTCGAGACACATGTTGTGTCACACTGGAAGAAAGACTTCTCTGTCTGATTTTGTGC 657
 Db |||||
 QY 903 TGTCGAGACACATGTTGTGTCACACTGGAAGAAAGACTTCTCTGTCTGATTTTGTGC 961
 Db |||||
 QY 658 ACAGA 662
 Db |||||
 QY 962 ARAGA 966

RESULT 13

AQ125617/c

LOCUS

DEFINITION

HS_2170 A2 H09 MF CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=2170 Col=18 Row=0, genomic survey

sequence.

ACCESSION

AQ125617

VERSION

AQ125617.1

GI:3502783

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 365)

AQ125617 365 bp DNA linear GSS 22-SEP-1998
 HS_2170 A2 H09 MF CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=2170 Col=18 Row=0, genomic survey
 sequence.

ACCESSION

AQ125617

VERSION

AQ125617.1

GI:3502783

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 365)

Qy	435	CAA	CAAGTGGTGA	CCTCACCTCTAAAGGTATGTTTGC	AAACTTTTGAAGACACGGGATGC	494
Db	380	CGCCAGCGGGCAT	CTCAGCTCAAGGTGTGCTCCAGACATTTGAAGATCCCAAGTCC	439		
Qy	495	TTCTGGAGCACCTTAAATCTCATG	CAGGCAAGTCTGTGTGGGGTTAAGAGAAAAAAGC	554		
Db	440	TGCTGGAGCACCTCAAGCTCATTTAG	CAGAGCTTCGTGGTGGCGAAGAAAGAAGC	499		
Qy	555	ACCAGTGGCAACATGTGATCG	CCGGTTCACACCGGAAGGATGTCGGAGACATGG	614		
Db	500	ATCCGTGTGACCATCGCGACAGAC	CTTCTACCCCGGAAGATGTGGGAGACACTTGG	559		
Qy	615	TGTTGCGACATGGGAAGAAGGACTTC	CTGTGAGTATTTGTCACAGAGATTTGGCGGNA	674		
Db	560	TAGTGCACACGGGGCGGAAGACTTC	CTGTGCCAGTACTGTGTCTCAGAGGTTTGGCCGGA	619		
Qy	675	AGGATCACCTGACTCGACATATGA	AGAAGTCAACATCAAGAGCTTCTGAGAGTCA--A	732		
Db	620	AGGATCATCTGACACGAGCACATGA	AGAAAGCCATCCACGGAACTGCTGGAACGATAACA	679		
Qy	733	AACAGAACCAAGTGGATTTCCTTTG	ACCCAT--TTACCTGCAATGTCTGTGCTTATAAAG	791		
Db	680	GACAGAGCCAGTTGACATGCTGG	GGTCTTTTAAGCTGCAGCTCGTCTGTAGCGGTACAAG	739		
Qy	792	ACGAGCTCCTTCC	804			
Db	740	AAGAGCTGAGTCC	752			

RESULT 15				
B99080				
LOCUS	B99080	386 bp	DNA	linear
DEFINITION	CIT-HSP-2283H1.TF CIT-HSP Homo sapiens genomic clone 2283H1, genomic survey sequence.			

LOCUS B99080 386 bp DNA linear GSS 26-JUN-1998
DEFINITION CIT-HSP-2283H1.TF CIT-HSP Homo sapiens genomic clone 2283H1,
genomic survey sequence.
ACCESSION B99080
VERSION B99080.1 GI:3026890
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 386)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shiruya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2283H1.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
7912 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

```

Class: BAC ends.
FEATURES             Location/Qualifiers
     1..386
         /organism="Homo sapiens"
         /mol_type="genomic DNA"
         /db_xref="GDB:7147972"
         /db_xref="taxon:9606"
         /clone="2283H1"
         /sex="Male"
         /cell_type="Sperm"
         /clone_lib="CIT-HSP"
         /note="vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII"

```

ORIGIN	Query Match	23.1%	Score 346.2	DB 8	Length 386
	Best Local Similarity	97.8%	Pred. No. 3.5e-90		
	Matches 351	Conservative 0	Mismatches 8	Indels 0	Gaps 0
QY	397	CAAGCTTTGGATT	TTAAACGTCACCTTGGCCCTTGCATCCGCGCAACAGTGGTGACCTCACCTG	456	
Db	28	CAAGCTTTGGATT	TTAAACGTCACCTTGGCCCTTGCATCCGCGCATCAAGTGGTGACCTCACCTG	87	
QY	457	TAAGGTAATGTT	TGCAAACTTTTGAAGACACGGGAGTGTCTTCTGGAGCACCTTAAATCTCA	516	
Db	88	TAAGGTAATGTT	TGCAAACTTTTGAAGACACGGGAGTGTCTTCTGGAGCACCTTAAATCTCA	147	
QY	517	TGACGGCAAGT	CGTGTCTGGTGGGTTTAAAGAAAAAAGCACACGCTGCGAACATTTGTGATCG	576	
Db	148	TGCAGGCAAGT	CGTGTATGTTGGTGGGTTTAAAGATTAAGACACACGCTGCGAACATTTGTGATCG	207	
QY	577	CCGGTTCTAC	CCGGAAGAGATTCGCGGACACATGGTGGTGACACACTGGAAAGAAGGA	636	
Db	208	CCGGTTCTAC	CCGCGAAAGGATGTCGCGGACACATGGTGGTGACACACTGGAAAGAAGGA	267	
QY	637	CTTCCCTCTG	TCAGTATTGTGCACAGAGATTGGGCGAAAGGATCACCTGACTCGACATAT	696	
Db	268	CTTACTCTG	TCAGTATTGTGCACAGAGATTGGGCGACAGGATCACCTGACTCGACATAT	327	
QY	697	GAAGAAGAGT	CAACATCAAGAGCTTCTGAAGGTCAAAACAGAACAGTGGATTTCTCTTG	755	
Db	328	GAAGAAGAGT	CAACATCAAGAGCTATTGAAGGTCAAAACATCAACAGTGGATTTCTCTTG	386	

Search completed: November 8, 2004, 20:41:39
Job time : 5077 secs

